Novel Strategies for Model-Building of G Protein-Coupled Receptors

Anske Stephanie van Neuren

Novel Strategies for Model-building of G Protein-Coupled Receptors.

Dissertation
zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat)

dem
Fachbereich Pharmazeutische Chemie
der PHILIPPS-UNIVERSITÄT MARBURG
vorgelegt von

Anske Stephanie van Neuren
aus Mijdrecht (Niederlande)

Marburg/Lahn 2003

Vom Fachbereich Pharmazeutische Chemie
der Philipps-Universität Marburg
als Dissertation angenommen am
19. August 2003
Erstgutachter: Prof. Dr.Gerhard KLEBE
Zweitgutachter: Prof. Dr. Thomas GUDERMANN
Tag der mündlichen Prüfung am: 19. August 2003

Die Untersuchungen zur vorliegende Arbeit wurden von November 1997 bis Januar 2001 in der Zentralen Forschung der Bayer AG, Ressort Life Sciences, Abteilung Molecular Modelling in Leverkusen und in der Arbeitsgruppe von Herrn Prof. Dr. Gerhard Klebe, Institut für Pharmazeutische Chemie des fachbereichs Pharmazie der Philipps-Universität Marburg unter Leitung von Herrn Dr. Gerhard Müller und Herrn Prof. Dr. Gerhard Klebe durchgeführt.

“Quis credat tantas operum sine numine moles ex minimis caecoque creatum foedere mundum? Si fors ista dedi nobis, fors ipsa gubernet. At cur dispositis vicibus consurgere signa et velut imperio praescriptos reddere cursus cernimus ac nullis properantibus ulla relinqui? Cur eadem aestivas exornant sidera noctes semper et hibernas eadem, certamque figuram quisque dies reddit mundo certamque relinquit?”

“ Who could believe that such massive structures have been created from tiny atoms without the operation of a divine will, and that the universe is the creature of a blind compact? If chance gave such a world to us, chance itself would govern it. Then why do we see the stars arise in regular succession and duly perform as at the word of command their appointed courses, none hurrying ahead, none left behind? Why are the summer nights and the nights of winter ever made beautiful with the selfsame stars? Why does each day of the year bring back to the sky a fixed pattern and a fixed pattern leave at its departure?”

(from the book *Astronomica* of the Roman poet Manilius, ˜ 20 Antum Domini)

“.. and why has a week 7 days and 7 nights, why are there 7 deadly sins and 7 wonders of the world, 7 oceans and 7 continents, why is Rome build on 7 hills, and “hast ein Buch 7 Siegel”, why has Little Tumb 7-mileboats, and Snow White 7 dwarfs and why has **an GPCR 7 transmembrane helices** ?
Just by chance ?”

Contents

Chapter 1

Scope of the Thesis

"I may disagree with what you have to say, but I shall defend, to the death, your right to say it."

- Voltaire
  (**1694-1778**), writer and philosopher)

Nearly every aspect of life on earth is controlled by processes that, from a biochemical point of view, rely on molecular recognition phenomena [1, 2]. The ability of individual molecules to bind selectively to structurally and functionally related target systems is a key principle for understanding biochemical reactivity, which affects processes such as enzyme catalysis, gene regulation, immunological defense, transmembrane signal transduction and numerous other important physiological events. The determination of molecular recognition phenomena in structural terms, especially those phenomena which encompass the interaction of small molecules (substrates, ligands) with highly complex biopolymers (enzymes, receptors), has been the aim of numerous research projects in medicinal, pharmaceutical and biophysical research. The attempt to rationalize the binding mode of, e.g., a low-molecular weight compound and a proteinogenic receptor system at an atomic resolution has developed into a vital component within the realm of structure-based drug design. More than 17000 experimentally determined high-resolution structures of proteins are currently known to the scientific community, numerous potential targets for drug discovery projects are among these. This situation is due to progress made over the last decade in, e.g., the development of high-technology instrumentation, such as synchrotrons and area detectors for X-ray crystallography, computer graphic systems connected to multi-processor supercomputer, and methodological developments in the area of high-field NMR spectroscopy. Despite this tremendous progress we have witnessed over the last decade, membrane-bound receptor systems are still underrepresented in the database of structurally characterized proteins, due to the problems associated with protein preparation, purification, and crystallisation or solubilization, respectively. Apart from the structures of an ion-channel [3], the $\alpha_v\beta_3$-integrin [4], and bovine rhodopsin [5], as a member of the GPCR family, not very many only a very small number of membrane-bound proteins relevant for drug discovery are currently determined at atomic resolution.

Consequently, the application of knowledge-based techniques from the realm of molecular modelling aimed at generating structural models for those proteins emerged as useful approaches towards a deeper understanding of the structural requirements underlying the corresponding molecular recognition events comprising experimentally non-amenable targets.

The research project presented in this study aims to introduce a novel approach towards the modelling receptor proteins from the target superfamily of G protein-coupled receptors (GPCR's), based on all principally available structure-relevant information originating from heterogeneous sources, such as site-directed mutagenesis, bioinformatics studies, structure-activity relationships of a series of low-molecular weight antagonists, protein modelling and molecular dynamics simulations. The relevance of those data for a model building approach is extensively evaluated throughout the different chapters of this thesis. The viability of the newly developed procedure is demonstrated with the generation, evaluation and validation of a complex between a model of the human CCK-B receptor and a series of experimentally characterized antagonists. The entire study is presented in five chapters. First is presented in chapter 2 an introduction on the basic principles underlying the signal transduction system of a GPCR. Thereby is emphasis laid on the current available information about GPCR structures. After this introduction an extensive study in terms of protein sequence analyses is presented in chapter 3 with the aim to unambiguously identify the transmembrane portions in the sequence of a GPCR, in other words, to generate the so-called 1D topology of the receptor. Chapter 4 describes the model building procedure exemplified with the human CCK-B receptor that, after molecular dynamics simulation under explicit treatment of the non-isotropic environment for energetic relaxation, revealed a receptor protein model that was pursued further to molecular docking studies. Prior

to a detailed description of the applied methodology for molecular docking and all associated comparative studies described in chapter 6, an overview is given on the various different chemotypes that qualified as promising lead series for CCK-A, as well as CCK-B antagonists (Chapter 5). Based on this overview, the entire study is designed along a path of increasing complexity, starting with protein sequence studies, followed by receptor protein modelling and subsequent simulations, finally yielding 3D models of molecular complexes comprising the target receptor and a series of low-molecular weight compounds. It is the intention of the author to emphasis that the entire modelling approach bears the potential to allow the experienced user to generate 3D models of any GPCR, provided a sufficient amount of structurally relevant input data is available.

Chapter 2

General Background

"GPCRs, such as the histamine H_2-receptor, can be regarded as proton pumps, which lack one piece in their proton shuttle to be able to pump continuously."

- P.H.J. Nederkoorn

  Thesis (**1996**)

2.1 G protein-coupled receptors

Among transmembrane proteins of current biomedical interest, the G protein-coupled receptors (GPCRs) constitute a superfamily of transmembrane receptor proteins that establish a functional and unidirectional link between the exterior of a cell and its cytoplasm [6, 7, 8, 9, 10]. A cascade of signal transduction events is induced by an extracellular occurring receptor-ligand recognition event and is transferred through the receptor by a yet unknown mechanism to the intracellular compartment involving binding to a heterotrimeric guanine nucleotide-binding regulatory protein: the so-called G protein.
The general idea is that the GPCR mediate the ligand action through a conformational change of the receptor, which is detected by the G protein [11, 8]. The conformation of the Gα subunit of the ternary complex changes its conformation so that the bound GDP molecule is displaced by an GTP molecule (the exchange mechanism). The Gα subunit dissociates subsequently from the G$\beta\gamma$-dimer. Both free subunits associate with effector enzymes, such as adenylate cyclase, guanylate cyclase, phospholipase A2 or C, stimulating or inhibiting production of second messengers including cAMP, cGMP, diacyl glycerol and IP3, which in turn cause downstream effects including the opening of Ca^{2+} or K^{+} channels and the generation of other messengers such as arachidonic and phosphatidic acid [8, 10] (Figure 2.1).

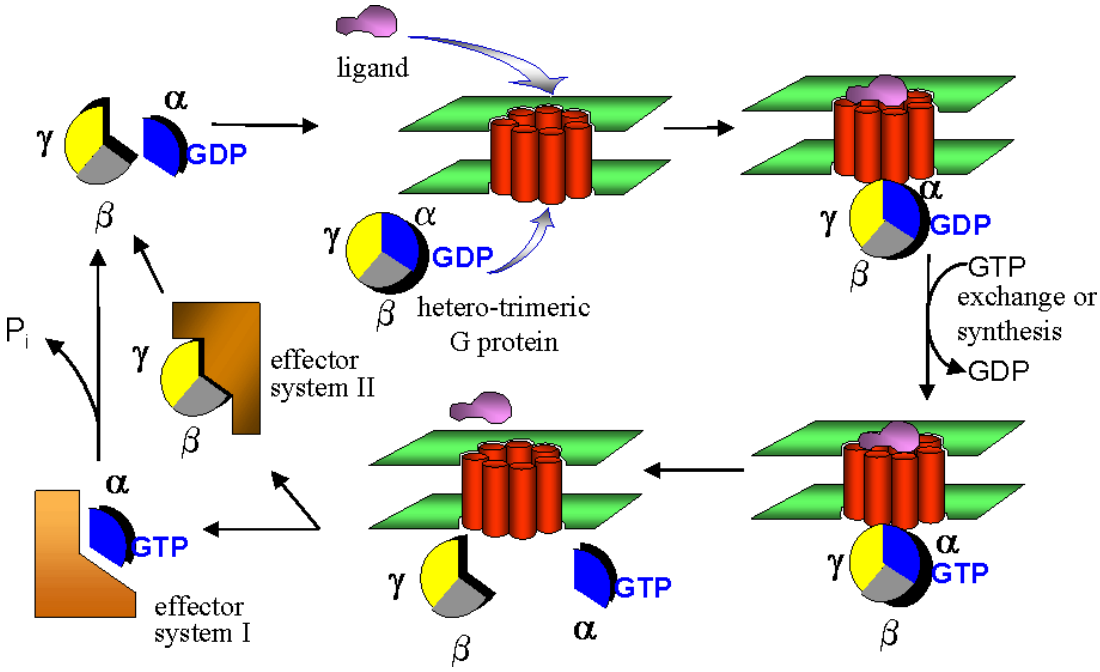


Figure 2.1: General signal transduction pathway for G protein-coupled receptors.

Several experiments [12, 13] published some years ago are not in accordance with the GDP-GTP-exchange mechanism. One would expect an increase in GDP concentration and an equal drop in GTP concentration upon activation. However, in one of those studies [12] it is reported that in rod outer segments, where rhodopsin and G proteins are abundant, GDP concentrations drop dramatically whereas GTP concentrations remain almost constant in the first 4 seconds after activation. Furthermore, it is reported that phosphate transfer occurs via Gβ subunits yielding GTP from GDP and P_i [14]. Such a phosphorylation reaction is ignored in the GDP-GTP-exchange mechanism, but this and aforementioned GDP concentration drop would

excellently agree with a GTP synthesis mechanism driven by a proton transport via the activated receptor. In a theoretical study of Nederkoorn et al. [7, 15] it was demonstrated that this ligand-mediated proton transfer is feasible for a series of GPCRs. The ligand-activation ends up with one or more protons per agonist bound, thus explaining why sometimes more than one G protein is activated by one bound agonist (signal amplification). In combination with the theory of Topiol [16], the deletion model for the origin of receptors, a GPCR can be regarded as a proton pump which lacks in the inactivated state one part of its proton pumping mechanism. An additional aspect is the similarity with the F_1F_0-ATP synthase [17, 18] and that the mechanisms of GTP synthesis and GDP hydrolysis (of the Gα subunit at the effector protein) resemble that of the phosphorylation and dephosphorylation reactions in nucleoside diphosphate kinases. Also other experiments (e.g. instability of uncomplexed Gα subunits [19], the double proton transfer of the tautomeric histamine molecule in the activation of the H_2-receptor [20], existence of constitutively activated mutant receptors [21], receptors with two different agonistic binding sites [22, 23]) are in agreement with this model and, in addition, this new model does not exclude some kind of conformational change of the GPCR [7, 15, and the references therein].

The proton-translocation-GTP-synthesis mechanism is not the only proposed "pumping" mechanism in GPCR signal transduction. Luca Turin [24] proposed that inelastic electron tunnelling would be the signal transduction mechanism for olfactory receptors, because not the molecular shape of the ligands but their vibrational properties are important for the activation of the receptor.

The multi-step cascade of signal transduction events is highly depending on the very nature of the distinct member of the GPCR family recognising its ligand with high specificity [25] (Figure 2.1). A broad spectrum of chemically diverse endogenous, as well as exogenous ligands, ranging from cations, biogenic monoamines, fragrances and taste molecules, lipidic mediators, oligopeptides of different size, up to large globular proteins elicit their biological activity by targeting their specific GPCR. Obviously, nature has designed a highly efficient transmembrane signal transduction system in that the great variety of receptor agonists converges towards the same type of target receptors, while the diversity is retained, since ligand-specific signalling cascades are triggered upon GPCR activation (Figure 2.2).

From a pharmaceutical research point of view, GPCRs represent one of the most prominent classes of validated drug targets. According to a survey from 1997, 25% of the one hundred world-wide top-selling drugs exert their therapeutic effect by targeting distinct receptors from the GPCR family [26], while the ratio referenced to all approved drugs is even 50 - 60% [27]. In terms of sales, GPCR-targeted drugs created a market potential that was estimated to approximately 16 000 million US $ during 1997 when referring to the top 100-selling drugs worldwide [26]!

A more detailed analysis of GPCRs, in terms of marketed drugs, reveals a strong bias in favour of the classical neurotransmitter receptors, such as the 5-HT, histamine, muscarinic acetylcholine, dopamine, or the noradrenaline receptor. Comparing the target-receptor distribution profile of released GPCR-targeted drugs with the number of identified receptors that can be attributed as therapeutically relevant target systems, a significant imbalance becomes immediately apparent, thus demanding future pharmaceutical research to focus on, e.g., peptide-binding GPCRs [28, 29].

Referring to the pharmacodynamic and -kinetic problems, generally associated with the chemical nature of peptides, modern medicinal chemistry will make extensive use of established peptidomimetic concepts to approach this goal following a ligand-guided strategy. A com-

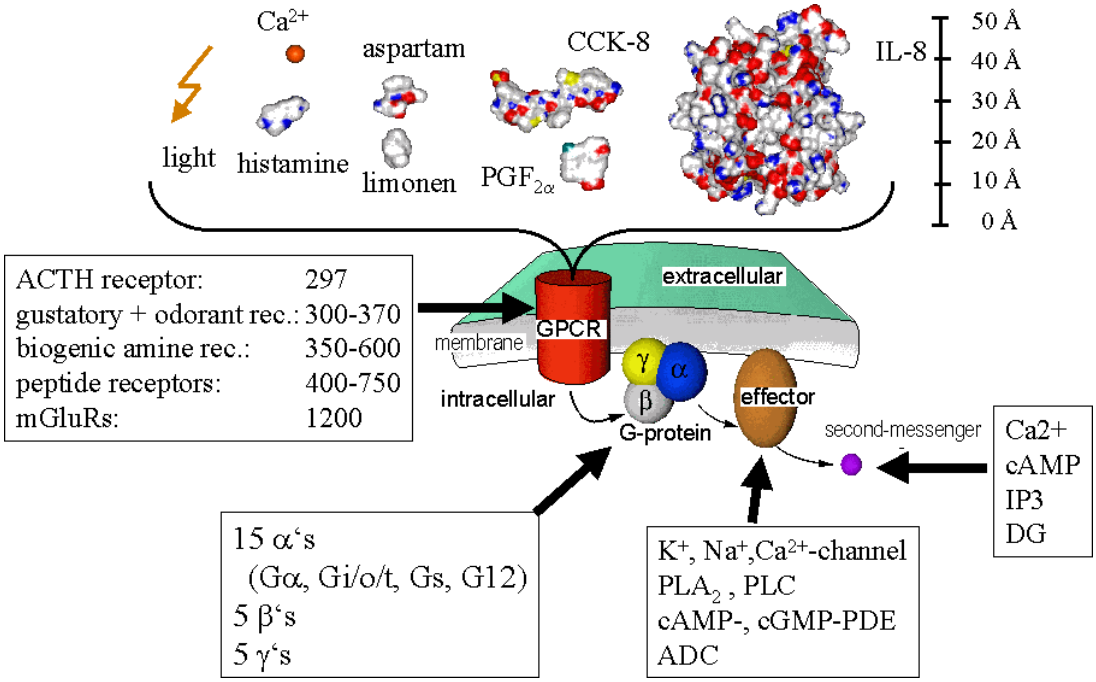


Figure 2.2: Diversity of ligands, sequences, G proteins and effectors versus functional convergence of G protein-coupled receptors.

plementary approach, relying on screening of random, diverse or targeted compound libraries, discovered already numerous non-peptide leads for peptide-binding GPCRs, interestingly most of them act as antagonists [30].

2.1.1 Binding Site Ambiguity

For the biogenic monoamines, a consensus binding mode within the transmembrane portion of the receptor protein was deduced. The putative binding site within the protein was mapped by mutational studies, while the key interactions, driving the corresponding molecular recognition event, were determined following classical comparative analyses on series of congeneric low-molecular weight compounds. These studies succeeded to correctly predict site-directed mutaganeses of the receptor protein or corresponding tailor-made modification of the ligand molecule, thereby supporting the structural hypothesis underlying an interaction model as shown in Figure 2.3 [31, 32]. Consequently, the numerous biogenic monoamine agonists and antagonists are accommodated in the same or highly overlapping binding pocket of in the target protein. They can thus be interpreted as classical competitive modulators.
As mentioned above, the comprehensive screening programs conducted over the last years mostly in pharmaceutical industry uncovered for almost any peptide-binding GPCR non-peptidic, low-molecular weight ligands, however the majority being antagonists [30].
From mutational mapping increasing evidence emerged that almost all peptide agonists interact with sequentially discontinuous epitopes on the extracellular surface of the transmembrane protein, while the non-peptide antagonists seem to address receptor residues further to the interior along the normal of the membrane surface [28]. The antagonists obviously penetrate the seven-helix bundle that all GPCRs are believed to be composed of, and they probably populate a common biogenic monoamine-related binding site (Figure 2.3).

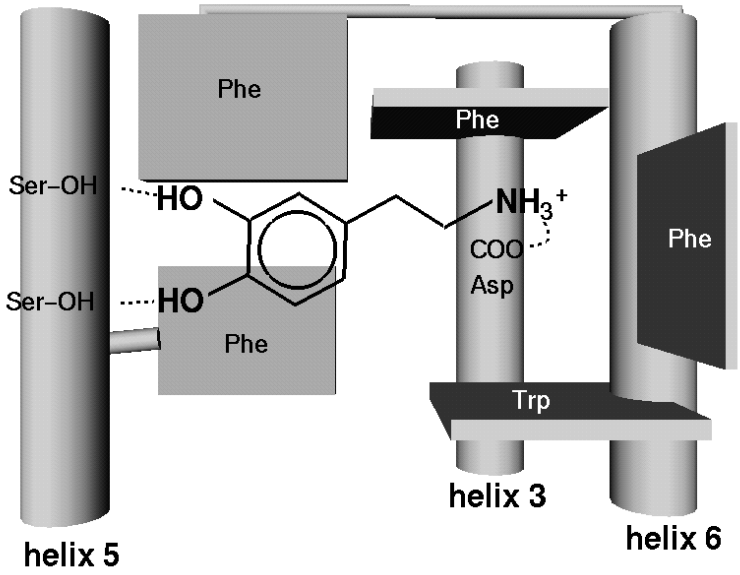


Figure 2.3: Schematic presentation of the consensus binding mode, identified for the biogenic monoamine-GPCR interaction. Transmembrane helices are depicted as vertical rods, aromatic rings of Phe and Trp sidechains are shown as squares. The ammonium ion of the ligands binds to an highly conserved Asp in helix 3 by an electrostatically driven interaction.

Consequently agonists and antagonists of peptide-binding GPCRs do not address the same "lock", even though some antagonists turned out to competitively displace the native ligand. The hypothesis of two spatially separated, not even overlapping binding sites for agonists and antagonists is further corroborated by the results of numerous diverse research projects on both molecular interaction partners, the receptor protein and the ligands, respectively. By constructing chimeric receptors it could be shown that the exchange of extracellular portions of the receptor protein allowed a directed shift of agonist-specificity, while non-peptide antagonist-binding was retained [28]. Even the complete deletion of a huge extracellular domain of, e.g., metabotropic glutamate receptors (mGluRs), proven to be responsible for agonist (glutamate) binding [33], yielded a receptor construct that could functionally be activated by non-peptide compounds [34]. In an elegant protein-engineering approach, a Zn^{2+} binding site was introduced into the neurokinin-1 receptor by replacing selected transmembrane residues against metal ion-coordinating histidines [28, 35]. These mutations did not affect native ligand binding, but spatially interferred with a non-peptide antagonist (CP-96,345) binding site, thus completely abolishing antagonist binding [28, 35]. Also a study of Ji et al. [36] on the angiotensin II receptor from *Xenopus laevis* underlined the spatially separated binding pockets of peptide agonist and non-peptide antagonist. The angiotensin II receptor from *Xenopus laevis* turned out to bind the native ligand with unaltered affinity, while the binding of the biphenyl-tetrazole drug Losertan could not be detected. Mutation of the 13 different transmembrane residues of the *Xenopus laevis* receptor with the human residues yielded a receptor capable of binding Losertan. This study and other available data clearly defeats the hypothesis of a common overlapping pharmacophore arrangement as is claimed for, e.g., biphenyl-tetrazole drugs and angiotensin II [37], whereby structural resemblance of the non-peptide antagonist and the corresponding peptide agonist was an important working hypothesis.

It should further be noted that, although a remarkable diversity is covered by the non-peptide antagonists of peptide-binding GPCRs, inherently reflecting the different origins of these compounds (in-house stocks, combinatorial chemistry-derived libraries, natural products), highly redundant structural elements can be identified among antagonists targeting different receptors [38]. These frequently recurring privileged substructures support a further idea, namely the existence of a binding pocket within the transmembrane domain, obviously highly conserved throughout the entire family of GPCRs that might correspond to the delineated binding site within the biogenic monoamine receptors [31, 32, 11] (Figure 2.3). This conserved binding site can be explained by the proton-translocation-GTP-synthesis mechanism [7, 15].
The hypotheses on binding sites and binding modes of GPCR ligands discussed above clearly demand further biochemical experiments and computer simulations that provide more detailed insights into the binding event of agonists and antagonists to their corresponding receptors. An improved 3D structural model could serve as a common working platform for "asking new questions" and assist in designing new experiments.

2.1.2 Structural Investigations on 7TM Proteins

Rhodopsins constitute a family of integral membrane proteins covalently binding a retinal molecule (vitamin A aldehyde) through a Schiff-base linkage to the -NH_2 group of a lysine residue that is located approximately in the middle between the cytoplasmic and extracellular surfaces of the protein. Members of these photoactive pigments fall into two distinct subfamilies, the prokaryotic and eukaryotic rhodopsins, respectively [39, 40]. Prokaryotic rhodopsins function as light-driven primary proton pumps (bacteriorhodopsin, archaerhodopsins), light-driven chloride ion pumps (halorhodopsin), and photosensors, mediating photoattractant as well as photophobic responses. In contrast, eukaryotic rhodopsins are G protein-coupled receptors acting as visual pigments in initiating the visual excitation cascade upon absorption of light (possibly via secondary proton-pump mechanism). The structurally unifying feature among all rhodopsins is a widely conserved protein topology characterised by seven sequentially consecutive transmembrane stretches that adopt α-helical conformations, linked by extracellular and intracellular loop regions with an extracellular N- and an intracellular C-terminus [9]. Despite of these common structural principles, the prokaryotic rhodopsins definitely do not belong to the family of GPCRs, as indicated by a low sequence homology between, e.g., bacteriorhodopsin and the visual pigments or other GPCRs of approximately 10% to 15% [9]. Thus, the 3D structure of prokaryotic rhodopsins such as bacteriorhodopsin or halorhodopsin cannot simply be accepted as a structural framework for molecular modelling attempts aiming at GPCR structural models.

2.1.2.1 Prokaryotic Rhodopsins

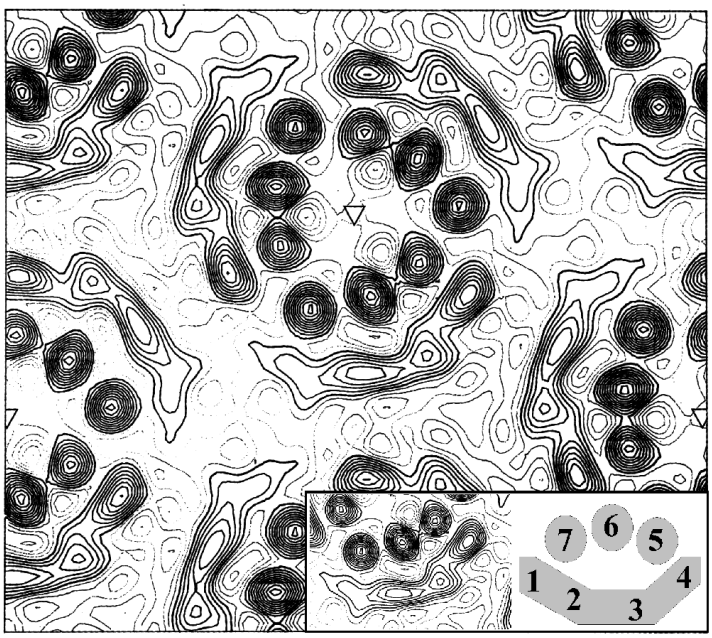
Undoubtedly, bacteriorhodopsin is by far the best characterised 7TM protein not only in terms of function, but also from a structural point of view. The fundamental observation that bacteriorhodopsin can be induced to form large and stable 2D crystals set the stage for all following structure determination efforts on 7TM proteins for many years. The first milestone was a 7 Å resolution map of the purple membrane obtained by Henderson and Unwin via electron microscopy and electron diffraction techniques in 1975 [41]. Even though this pioneering work dates back 27 years, the first 3D structure of bacteriorhodopsin with a resolution comparable to

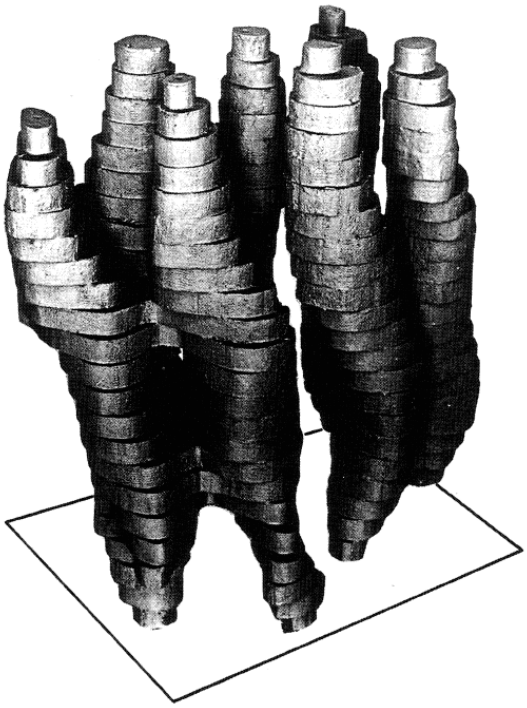
that usually obtained for soluble proteins was described just in 1997 [42]. This breakthrough became possible through a novel crystallisation strategy yielding 3D crystals of sufficient diffraction quality for synchrotron studies [43].

2.1.2.2 2D Crystals of Bacteriorhodopsin

Since 3D crystallisation followed by crystal structure determination were not successful until 1999, the vast majority of structural studies on 7TM proteins applied electron microscopy in combination with computational image processing of 2D crystals. A 2D crystal is a highly ordered, single-layered crystalline specimen of very limited thickness (with vertical unit-cell dimension), obtained by detergent-mediated reconstitution of solubilised and purified protein into bilayers, closely resembling a native, membrane-like environment [44, 45, 46, 47]. Within these planar lattices that occur as vesicles, sheets, or tubes, repetitive subunits of identical composition appear in equivalent positions with a high degree of symmetry, being the precondition for extracting information about structural details by electron microscopy and diffraction techniques [44, 45, 46, 47]. Although the information on structural details is not visible on the primary image, computational processing in combination with electron diffraction patterns help to elucidate structural details close to atomic resolution. In practise, a series of projections obtained at different tilt angles are combined to reproduce a 3D image of the investigated object that shows high resolution in the in-plane dimensions and decreased resolution in the vertical direction. Compared to single crystal diffraction techniques, electron microscopy-based imaging and diffraction collects deviating information. In X-ray crystallography the intensity of the diffracted beam can be related to the atomic scattering factors. In electron diffraction the relationship to atomic scattering factors is more complex. The electrons interact with the Coulombic potential that arises from the negatively charged electron cloud, as well as from the positively charged nuclei. Therefore, electron form factors (atomic electron scattering cross-sections) are more sensitive to the physicochemistry of atoms, such as the nature of chemical bonding, ionised functionalities, or polarised, partially charged groups. One advantage of electron crystallographic structure determination over X-ray crystallography is that phase information, often a major obstacle in X-ray-based structure determination, is available from images taken with the electron microscope [44, 45, 46, 47].
The 3D map of Bacteriorhodopsin was described as dominated by numerous "rod-shaped features" aligned perpendicular to the membrane surface [41]. Within each asymmetric unit seven rods form a molecule with overall dimensions of 25 Å x 35 Å x 40 Å (X x Y x Z, Z being perpendicular to the membrane plane). All rods span a distance of 35 Å to 40 Å along their main axis, perfectly corresponding to the thickness of the 2D crystal. Each protein molecule contributes to a trimeric bundle with a threefold symmetry axis [41]. As first established in 1975, the rods were assumed to represent the seven transmembrane α-helices forming the 7TM domain of retinal proteins.
The structure model, released as 1BRD to the Brookhaven Protein Data Bank was based on the potential map derived from 72 images of tilted specimen recorded at the temperature of liquid helium [48]. Combining electron cryo-microscopically derived micrographs and electron diffraction patterns allowed to drive the resolution of the resulting projection map up to 3.5 Å in the membrane plane and 7.8 Å perpendicular to the membrane surface [48].
It is noteworthy that in this study the retinal molecule was clearly visible. The chromophore is covalently bound to the -NH_2 group of Lys-216 and governs the gating of the proton pump by

A





B

Figure 2.4: (A) Projection of the map of electron scattering density of bacteriorhodopsin onto the plane of the membrane at 7 Å resolution [41]. The individual molecules are arranged about a 3-fold rotational axes of symmetry normal to the plane of the membrane (triangles) to form trimers in the membrane. The projection of one molecule with designed helix-numbering is depicted in inset. (B) A 3D-reconstruction of bacteriorhodopsin made of sections through the 3D map of electron scattering density at 2 Å interval and traced onto balsa wood [41].

its isomerisational switch. The Schiff-base forms a kind of bottom of two half-channels protruding from the centre of the helix bundle towards the extracellular and intracellular surface, respectively. The residues delineating both half-channels could clearly be identified from the experimental data and were further assumed to contribute specifically to the molecular mechanism of the ion transport process [48].
By expanding the available experimental data set by 30 additional images of tilted specimen, thus contributing phase information to the density map, the resolution could further be enhanced from 7.8 Å (1BRD) to 4.3 Å (2BRD) in the vertical dimension [49] (Figure 2.4). Apart from the first extracellular and the first and third intracellular loops which show no interpretable density, the entire protein is visible. Interesting structural details became obvious with this refined structure which were undetected or even misinterpreted in the precursor structure. In this context, transmembrane helix 4 had to be altered with respect to the initially released model by a 4 Å shift towards the cytoplasm. Further it turned out that transmembrane helices 2, 3, and 6 show an overall kinked shape due to proline residues near the centre of the helical stretches (Pro-50, Pro-91, Pro-186) [49].
A further high-resolution electron crystallography study on bacteriorhodopsin appeared in 1997 [50]. The derived data even allowed to improve the in-plane resolution further up to 3.0 Å (1AT9), thereby revealing new structural details, especially an improved surface structure of the transmembrane protein. While in 2BRD three loops were assigned as disordered, in the latest study of a Japanese group all loops are shown to be highly structured [50]. With the refinement of the water-accessible surface area of the transmembrane protein, new features, probably driving the transport mechanism, became apparent.
Apart from studies on bacteriorhodopsin, much effort has been spent on the structure elucidation of the light-driven chloride-ion pump halorhodopsin from *Halobium salinarium*. As net result from a series of investigations, the projected structure of halorhodopsin seems to be almost identical to that derived for bacteriorhodopsin, thus reflecting the close sequential relation between these proteins [51, 52].

2.1.2.3 X-ray Structure of Bacteriorhodopsin

At the end of 1996, Landau and Rosenbusch published their promising results on obtaining 3D mircrocrystals of bacteriorhodopsin [53, 42] utilising a new concept of membrane protein crystallisation from 3D membrane-mimetic matrices, afforded by a distinct morphological state of lipid membranes, notably the lipidic cubic phases [54, 55]. The major refinement of the bacteriorhodopsin structure achieved by the X-ray study clearly refers to the enhanced spatial resolution along the main helix axis of the seven α-helices constituting the 7TM domain. The 2.5 Å structure closely resembles the structural details elaborated over two decades by electron microscopy, image reconstruction, and electron crystallography on 2D crystals. However, distinct differences can be found in loop conformations and in the orientation of sidechains from selected residues. Due to the well resolved positions of functionally important amino acid sidechains, together with the precise position of interacting water molecules presumably participating in the proton transport, a high resolution picture of a likely proton pathway emerged from that study [53, 42].
Only 9 months later the X-ray structure of bacteriorhodopsin was published [42], a further structure with slightly improved resolution (2.3 Å) was solved by taking advantage of the very same crystallisation technology employing cubic phases [56]. In this study, the authors observed an

unexpected C2 symmetry, presumably due to twinned crystals, which was explicitly accounted for within the refinement strategy. Apart from a different number of localised water molecules aligning the proton-conducting pathway, the overall structure reproduces the findings derived previously. As in 1AP9 [42], the newly released structure 1BRX [56] does not show any interpretable density for the N- (1-5) and C-termini (229-248), respectively. Also parts of the third cytoplasmic loop (154-166) are obviously disordered.

2.1.2.4 Eukaryotic Rhodopsins

Unlike the rhodopsins from bacterial sources, the eukaryotic rhodopsins belong to the protein superfamily of GPCRs, thus qualifying them as ideal candidates to derive experimental structural data indicative for this target family, since these proteins also tend to form 2D crystals amenable to electron microscopy and crystallography methods. In contrast to the bacterial rhodopsins that function as primary ion pumps, eukaryotic rhodopsins are visual pigments. They trigger the aforementioned GPCR-G protein signal transduction cascade upon light absorption, thereby inducing membrane hyperpolarization and nerve excitation [39, 40]. Comparative protein sequence analyses reveal amino acid sequence homology of eukaryotic rhodopsins with several members of the GPCR family, supporting not only the functional relation between rhodopsins and GPCRs, but also suggesting that the numerous members of this pharmacologically interesting target family are structurally related to the rhodopsins.
A huge body of structurally relevant data were derived over the last few years based on electron cryo-microscopy, image processing, and electron crystallographic studies of 2D crystals with steadily increasing spatial resolution. Projection structures of bovine rhodopsin [57], frog rhodopsin [58], and squid rhodopsin [59] were calculated at resolutions that clearly allow to define the topological differences between GPCRs on one hand and bacterio- and halorhodopsin on the other hand. This was confirmed in 2000 by the X-ray structure of bovine rhodopsin [5].
The history of structural elucidation of eukaryotic rhodopsins is different from that of bacteriorhodopsin in that it is characterised by a steady interplay between experimental structural studies on receptor proteins from different vertebrate and invertebrate sources. Even though no high resolution structure was available until the year 2000 [5], a prototype 3D structure model for the 7TM domain of the rhodopsin family of GPCRs could be deduced from an iterative structure refinement procedure, permanently interfacing structural bioinformatics studies conducted by Joyce Baldwin [60, 6] into the structure reconstruction attempts. With respect to the X-ray structure determined in the year 2000 [61], the model generated by Baldwin can be ranked as “very reasonable”.
The first projection structure of bovine rhodopsin with an in-plane resolution of 9Å was published in 1993 [57]. The resulting projection density map revealed interesting details about the arrangement of the transmembrane helices. The density map displays an arc-shaped feature in proximity to four resolved peaks of density (see first picture of Figure 2.8). While the arc-shaped density was assigned to three tilted or kinked helices with overlapping termini in the projection plane, the four resolved peaks were interpreted as four helices oriented perpendicularly in the membrane. Since no structural data were collected from tilted specimen, only 2D structural information in terms of the projection map emerged from this investigation [57].

2.1.3 Structural Data Meet Sequence Data

Two weeks after the manuscript on the projection structure of bovine rhodopsin was submitted for publication, Joyce Baldwin submitted a paper on the tentative structure of the transmembrane domain of GPCRs [60]. Structural implications for the 7TM domain are extracted from a thorough comparative sequence analysis study and combined in an elegant way with the experimentally derived structural data on bovine rhodopsin discussed above [57]. Distinct features in the amino acid sequence of GPCRs were analysed from alignments of approximately 200 GPCR sequences [60]. From this multiple alignment strategy, a "fingerprint" was deduced identifying highly conserved patterns in the primary structure of the 7TM domain (Figure 2.5). Based on

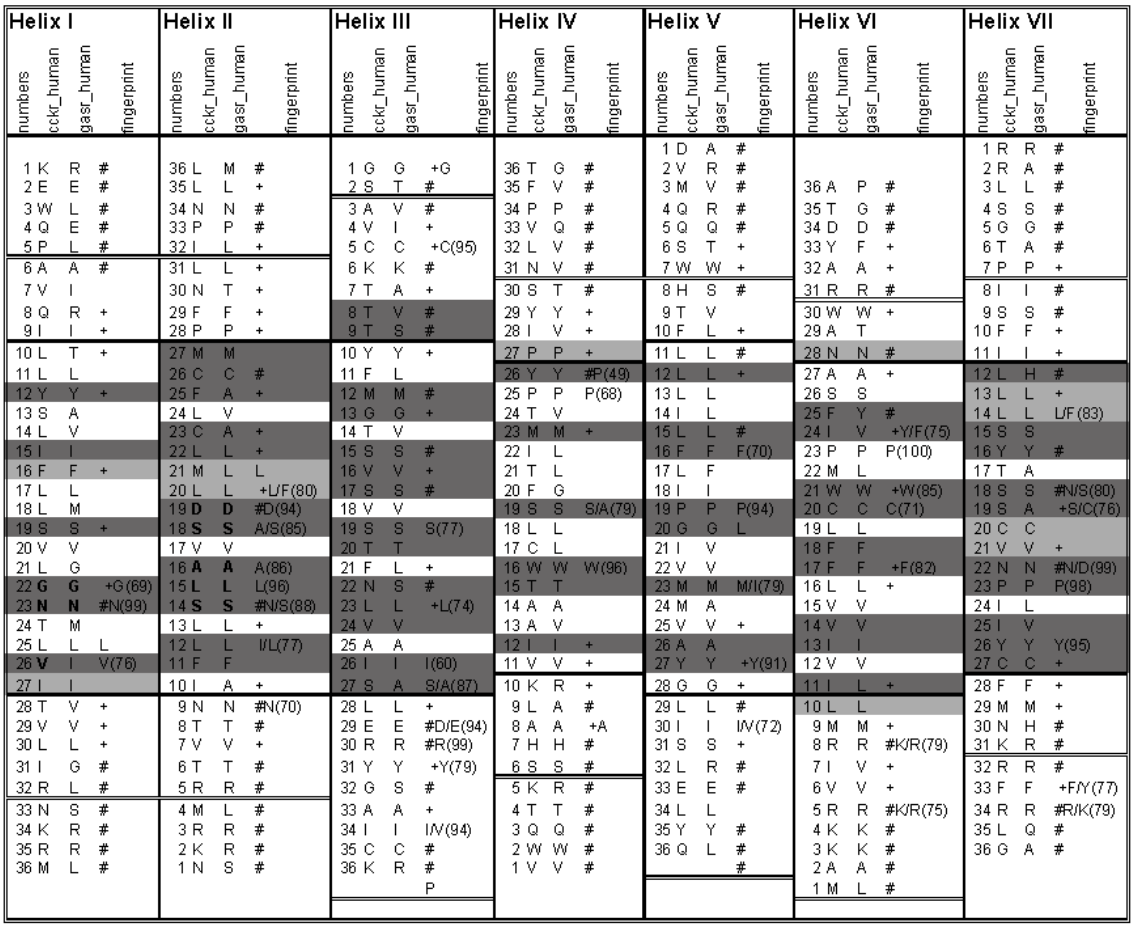


Figure 2.5: The fingerprint derived by Baldwin for the 7TM domains of approximately 200 GPCR sequences

this sequence pattern, the putative transmembrane sequence stretches spanning the hydrophobic core of the bilayer membrane were assigned under the following assumptions:

- the 7TM domain is formed by ideal α-helices, each consisting of 26 residues;
- 18 residues are required to span the hydrophobic core of the membrane, no polar or charged residues are expected to face the lipophilic environment in this part of the sequence;

- four residues flank the 18 core residues on the extracellular and the cytoplasmic surface, respectively.

These qualitative, 1D structural data were expanded by 2D structure information by imposing the fingerprint onto helical wheels. This approach revealed helices 1, 4, 5, and 6 to display large surfaces with nonpolar residues, assigning them to be significantly exposed to the lipophilic environment. Helices 2, 3, and 7 seem to be more buried, since their hydrophobic radial segments are definitely smaller [60].
A more detailed consideration of sequence characteristics allowed for a clear discrimination of lipid-facing surface patches from helix contact areas and from the interior of the helix bundle. The assignment of residues oriented inwards into the helix bundle was corroborated by the consideration of residues, identified to be involved in ligand binding from numerous mutagenesis studies. Finally, the distinct helical wheels were assembled to a 7TM helix bundle, thus representing cross-sections of the receptor protein taken at different heights along the normal of the membrane. Helices 1, 4, and 5 were found to be exposed to the lipophilic exterior, while helix 3 turned out to be well protected from the membrane environment. Additionally, the anticlockwise arrangement when viewed from the extracellular side (outside-in view) emerged as the most likely circular topology for GPCRs, since only this arrangement accounted for the majority of available mutagenesis and ligand binding data on GPCRs [60].
Thus, valuable structural information could be deduced from a sequence analysis on approximately 200 GPCR sequences that yielded not only a rough assignment of the transmembrane sequence stretches, but also resulted in a 3D model, the characteristics of which turned out to be compatible with the low-resolution projection map of bovine rhodopsin [57].

2.1.4 Enhanced Resolution of Rhodopsin

In 1995 Schertler et al. succeeded in improving the structural details of GPCRs, by combining micrographs taken from 2D crystals with different tilt angles [62]. After image processing a 3D structure could be obtained, still with poor vertical resolution. However, from the cryomicroscopically derived 3D structure, four clearly resolved helices together with three remarkably tilted helices could be identified, clearly resembling the 3D model obtained from the GPCR sequence analysis discussed above [62].
In the year 1998 crystals from bovine rhodopsin were obtained that were sufficient in size and order for structure elucidation by means of electron crystallography [64]. The horizontal resolution could be improved to 5 Å. Helices 4, 6, and 7 are visible as isolated density peaks and seem to be inserted almost perpendicular into the membrane, while helices 1,2, and 3 are tilted and overlap in the projection. In contrast to the previously described low resolution structure [62, 57], helix 5 significantly deviates from a perpendicular orientation [64].
Not only structural data on bovine rhodopsin [64, 62, 57] were accumulated, but also data on other vertebrate and invertebrate rhodopsins [58, 59, 62, 65]. The results obtained from microscopical and diffraction techniques on frog rhodopsin [58, 62] complement the work on the bovine receptor protein [64, 62, 57]. Given a sequence identity of 85% between both proteins, bovine and frog rhodopsin are assumed to share very similar, if not identical 3D structural features in the transmembrane domain. From the enhanced resolution structure of frog rhodopsin it became obvious that helix 5 was more tilted than anticipated from the very first study on the bovine protein. The density map at a resolution of 6 Å shows three resolved and isolated peaks

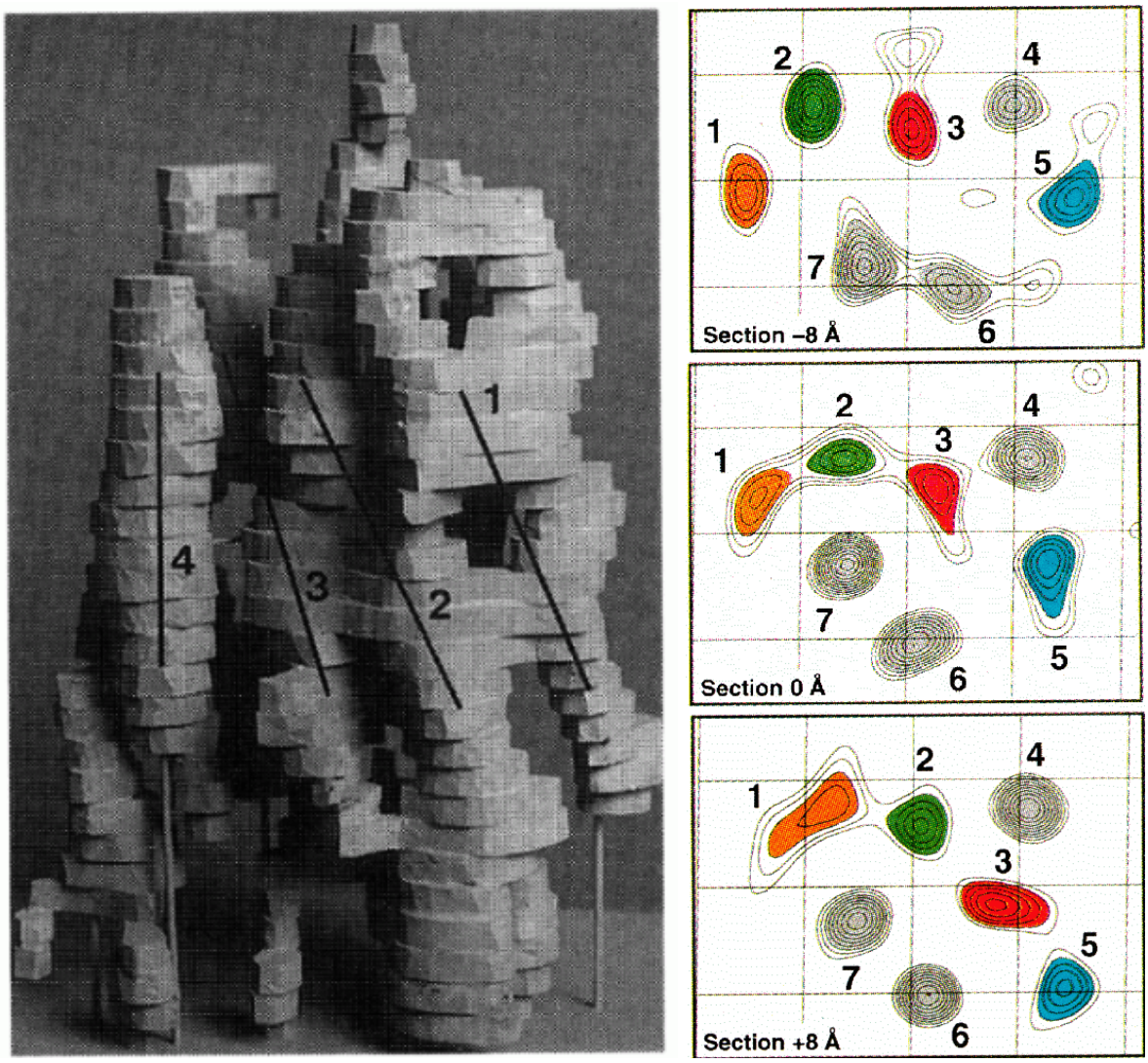


Figure 2.6: Schematic presentation of the reconstructed frog rhodopsin topology (left), derived from a stack of density maps taken from different heights along the membrane normal. The density maps (right) are shown in an inside-out view, revealing the clockwise arrangement of the seven transmembrane helices. The pictures and a detailed description of the biophysical studies on rhodopsins can be found on the internet pages from the Schertler group at the MRC (http://-www2.mrc-lmb.cam.ac.uk/groups/GS/rhodopsin.html) or in the article of Unger and Schertler [63].

in a triangular arrangement, accounting for helices 4, 6, and 7 and a contiguous band of density extending through the centre of the molecule. This density feature arises from a series of tilted helices that are not sequentially arranged in a consecutive manner. For the final helix assignment, the authors interfaced the results from the exhaustive structural bioinformatics study from J. Baldwin [60, 6]. The most interesting finding from this study refers to the orientation of helix 5, which turns out to be in contact with helix 3 due to the pronounced inclination angle of the helix axis [58]).
In 1997, a further improved 3D structure of frog rhodopsin was reported [62] (Figure 2.6). The achieved resolution allowed to compute approximate tilt angles for all seven helices. The tilt angles for the three perpendicular helices 4, 6, and 7 range from 3.8° to 13.4°, while helices 1, 2, 3, and 5 display values between 22.7° and 30.0°, respectively. Due to the significant tilted insertion mode of four helices, the specific helix-helix interactions are not restricted to sequentially adjacent transmembrane sequence stretches. Also the degree of lipid exposure of distinct helices changes significantly along their pathways through the lipid bilayer.
The most tilted helix 3 is deeply buried in the protein, its cytoplasmic end being surrounded by helices 2, 4, 5, and 7. On its way to the extracellular surface the helix axis drifts away from helix 5. In the centre of the membrane, helix 3 is in close proximity to helix 7 establishing a "through-bundle" contact. Helix 4 seems to be the shortest and least tilted helix.
The overall helix packing is more dense on the cytoplasmic surface and the helix bundle opens up ascending towards the extracellular side, thereby forming a cavity delineated by surface patches from helices 3, 4, 5, 6, and 7 accessible from the extracellular compartment. This is undoubtedly interesting for the entire family of GPCRs in terms of ligand binding capabilities [62].

2.1.5 3D Structure Template for GPCRs

Baldwin expanded her study to approximately 500 GPCR sequences and included further indirect structural data obtained from a series of biochemical studies on numerous GPCRs elaborated by different research groups [6]. It was the frog rhodopsin density map with considerably high resolution vertical to the membrane plane [58] which Baldwin et al. [66] used to generate a 3D model for a prototype GPCR (Figure 2.7). Based on the density maxima observable for all seven helices of frog rhodopsin in the corresponding projection map over a 20 Å axis perpendicular to the membrane plane, the trace of each helix across the lipid bilayer could be reconstructed and so a first atomic model with Cα-positions of each transmembrane amino acid adopting a helical conformation could be constructed including helix kinks induced by proline residues [66]. For helices 5 and 6 the occurrence of prolines in the transmembrane sequence stretches produces significant changes of the inclination angle relative to the normal of the membrane plane.
According to the rules for transmembrane helix bundles referring to amphiphilicity, as well as conserved and variable sequence positions [67, 68], the lipid-facing surface patches could be unambiguously identified. Together with the cross sections of the frog rhodopsin density map, the orientation of each helix with respect to the centre of the bundle was clearly deducible.
Helices 1, 4, and 5 exhibit large lipid-facing surface areas displaying large sections of variable sequence positions. A high degree of conservation, being an indicator for structural or functional important residues, was identified for large surface areas of helices 2, 3, and 7. From the model building procedure it became obvious that the membrane-water boundary not necessarily

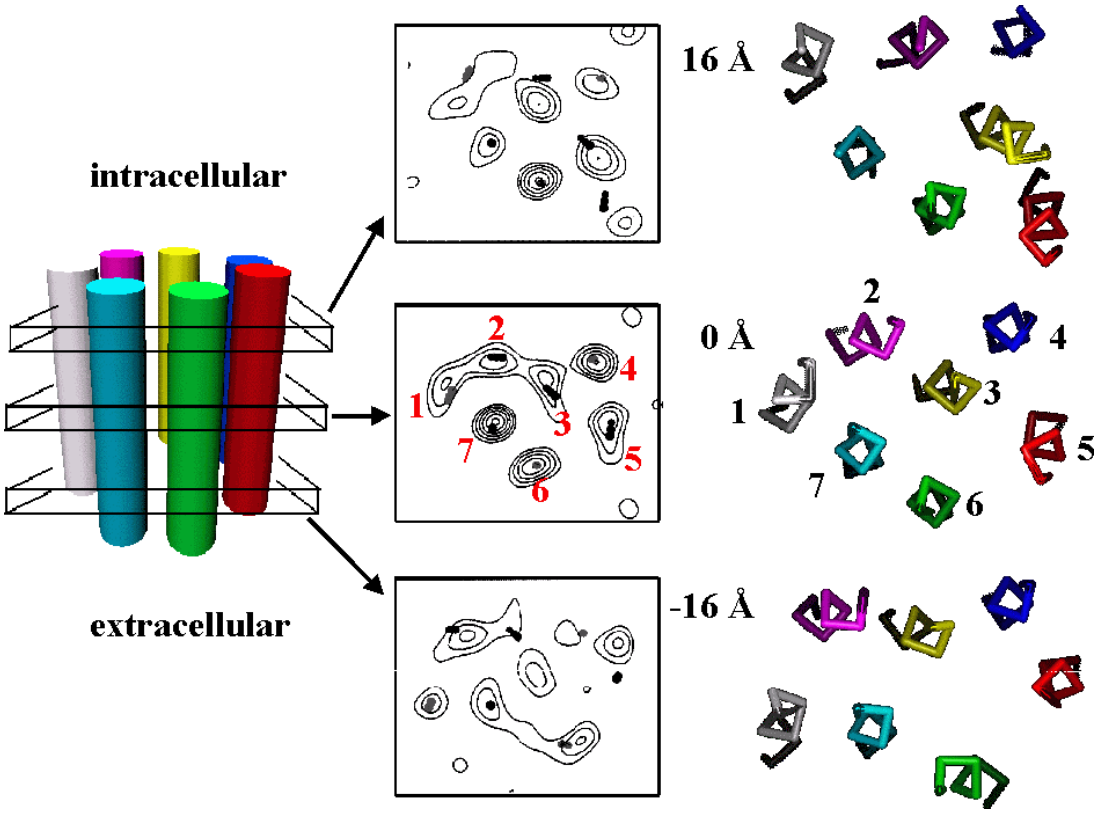


Figure 2.7: In-plane cross sections (right column) of the Cα-template (coordinates were kindly provided by Joyce Baldwin) taken over a transmembrane range of 32 Å. The 0 Å plane defines the central core section of the membrane, while the upper and lower picture represent the cytoplasmic and extracellular interface region, respectively. The corresponding density maps are shown in the center.

disrupts the helical conformation of the transmembrane sequences. Although an uncertainty remains for the assignment of the exact helix termini, some of them clearly protrude from the membrane surface into the flanking aqueous compartments [66]. The final structural model was extensively verified against results from various protein engineering studies on different peptide-binding GPCRs [66, and references therein].
The differences between GPCRs and bacteriorhodopsins described in the next section will demonstrate that any modelling attempt of GPCR structures employing the bacterial receptor protein structure remain highly questionable.

2.1.6 Structure Scaffolds for GPCR-models: Bacteriorhodopsin versus Rhodopsin

The majority of GPCR modelling attempts described in literature aimed to obtain 3D structural models of the 7TM domain for rationalising, e.g., structure-activity relationships of low-molecular weight agonists and/or antagonists within the putative ligand binding pocket of the target receptor employed the coordinates of bacteriorhodopsin 1BRD as underlying structural scaffold [32]. This is certainly due to the availability of this coordinate set to the scientific community via the Brookhaven Protein Data Bank since 1990 [48]. In light of the structural

and functional knowledge on rhodopsins derived over the last years, rhodopsin structures from bacterial sources can no longer serve as templates to built GPCR structure models on.
On the level of primary structure, no significant homology can be detected between bacteriorhodopsin and, e.g., human rhodopsin. A sequence similarity search over protein sequence databases with vertebrate rhodopsins as queries unravels numerous sequences of GPCRs to be homologous with significant sequence identity to the query. Using bacteriorhodopsin as query, *almost* no GPCR sequences are found with high similarity or identity scores.
Although no structure of a vertebrate rhodopsin was determined at a comparably high resolution as for bacteriorhodopsin (until 2000 [1]), several structural features were identified to be entirely different in both protein classes. In contrast to bacteriorhodopsin, which has no cysteine residues, almost all eukaryotic GPCRs form a highly conserved disulphide bond between the extracellular terminus of helix 3 and the second extracellular loop connecting helices 4 and 5 [9]. A comparable structural constraint is totally absent in bacteriorhodopsin. In general, membrane cross sections of the vertebrate rhodopsin structures are described to appear as compact entities, while the bacterial receptors exhibit an elongated helix bundle when projected into the membrane plane [63, 65] (Figure 2.8). Also the pattern of tilt angles is significantly different

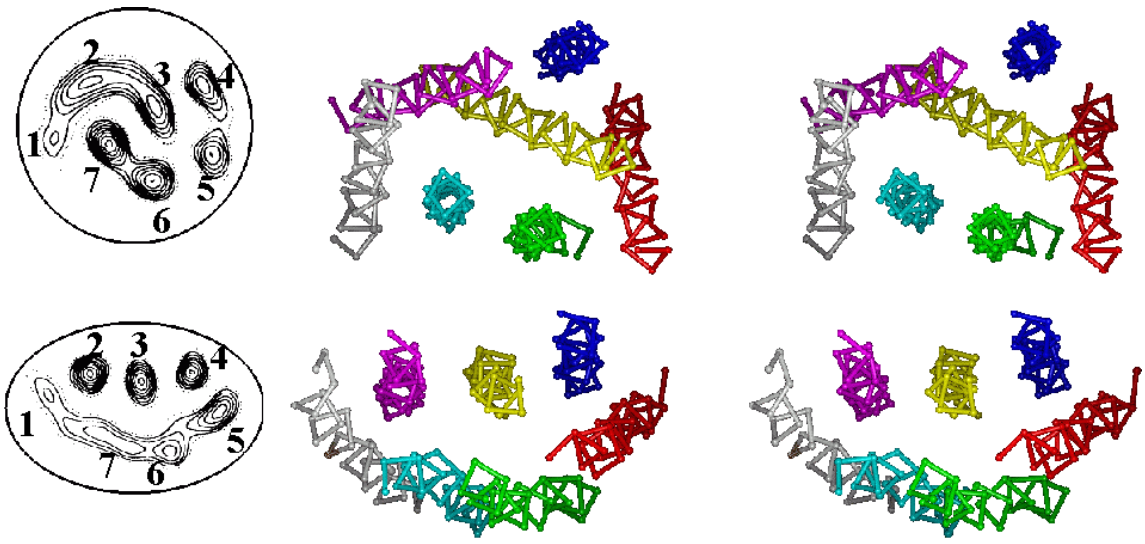


Figure 2.8: Left: Experimentally derived electron scattering projection maps of bovine rhodopsin (above) and bacteriorhodopsin (below). Right: Side-by-side stereo picture of the Cα-trace-model of vertebrate rhodopsins compared to the X-ray-derived structure of bacteriorhodopsin (PDB entry code: 1AP9). The density maps, as well as the helix bundles are shown in an inside-out view.

when the projection map of bacteriorhodopsin is compared to either frog [62, 58], squid [59], or bovine [57, 63, 64] rhodopsin. In the bacterial helix bundle, helices 2, 3, and 4 show an almost perpendicular arrangement, while the remaining helices overlap to form an extended arc-shaped density trace [49]. In contrast, helices 4, 6, and 7 are oriented along the membrane

[1] In 2000 the X-ray crystal structure of bovine rhodopsin was published. This was after all the modelling studies in this thesis were prerformed. Therefore this structure could regrettable not be taken as scaffold structure in the modelling studies, but in the final notes a comparison between this high resolution X-ray structure of rhodopsin and our modelled structure of the human CCK-B receptor has be made (see Chapter 4).

normal in the eukaryotic receptor proteins, while helices 1, 2, 3, and 5 partially overlap in an in-plane projection (Figure 2.8). Also the distribution of helix-distorting proline residues in the transmembrane helices shows an entirely unique pattern. Prolines are found in helices 2, 3, and 6 in the bacterial receptor, and in helices 4, 5, 6, and 7 in the vertebrate rhodopsins. The combination of the different membrane insertion modes and a different pattern of helix kinks clearly reveal the substantial differences of both receptor proteins.
Functionally, the bacterial proteins act as light-driven primary ion pumps [42], facilitating transport of matter through the transmembrane helix bundle against a proton gradient (up-hill), while eukaryotic rhodopsins work as switching moieties, possibly by a down-hill proton translocation mechanism, transmitting a signal across the membrane [7, 15]. This functional difference is further reflected by the different ground states of the covalently bound cofactor, retinal. In bacteriorhodopsin the absorption of light triggers isomerisation of all-trans retinal to the 11-cis isomer, while in vertebrate rhodopsin the 11-cis isomer represents the ground state.
In the light of these pronounced structural differences between the ion-pump bacteriorhodopsin and the vertebrate rhodopsin GPCRs, one should refrain from employing any bacteriorhodopsin structure as underlying framework for GPCR modelling purposes.
Structural models of GPCRs based on the current state of available experimental input data can by no means be compared to classical homology models of soluble proteins. They clearly lack the degree of spatial resolution and structural refinement necessary for giving an approximately correct picture of the atomic details that would be mandatory for, e.g., structure-based drug design or the derivation of quantitative estimates of ligand-receptor interactions. However, GPCR models have to be seen as a very useful working platform to discuss findings from molecular biology and medicinal chemistry and to create new ideas about possible experiments.

2.1.7 Crystal Structure of Rhodopsin

In August 2000, after the modelling study described in this thesis was accomplished, the very first crystal structure at 2.8 Å resolution of bovine rhodopsin was published by Palczewski et al. [5]. It shows the seven transmembrane domains, the 70 residues of the cytoplasmic domain - the three intercellular loops and the carboxyl terminus - and the 74 residues of the extracellular region (Figure 2.9). In the same issue of Science, Bourne and Meng [61] described the superposition between the crystal structure of rhodopsin and three recent model structures and found that the Cα-trace of the transmembrane helices fitted reasonably well (RMS-deviations of 3.1 - 3.2 Å). Of course from now on, it is preferable to use this structure of rhodopsin [5] as template for any modelling study of a GPCR, but also former models based on the mentioned assumptions together with structurally relevant data resulted in reasonable models.

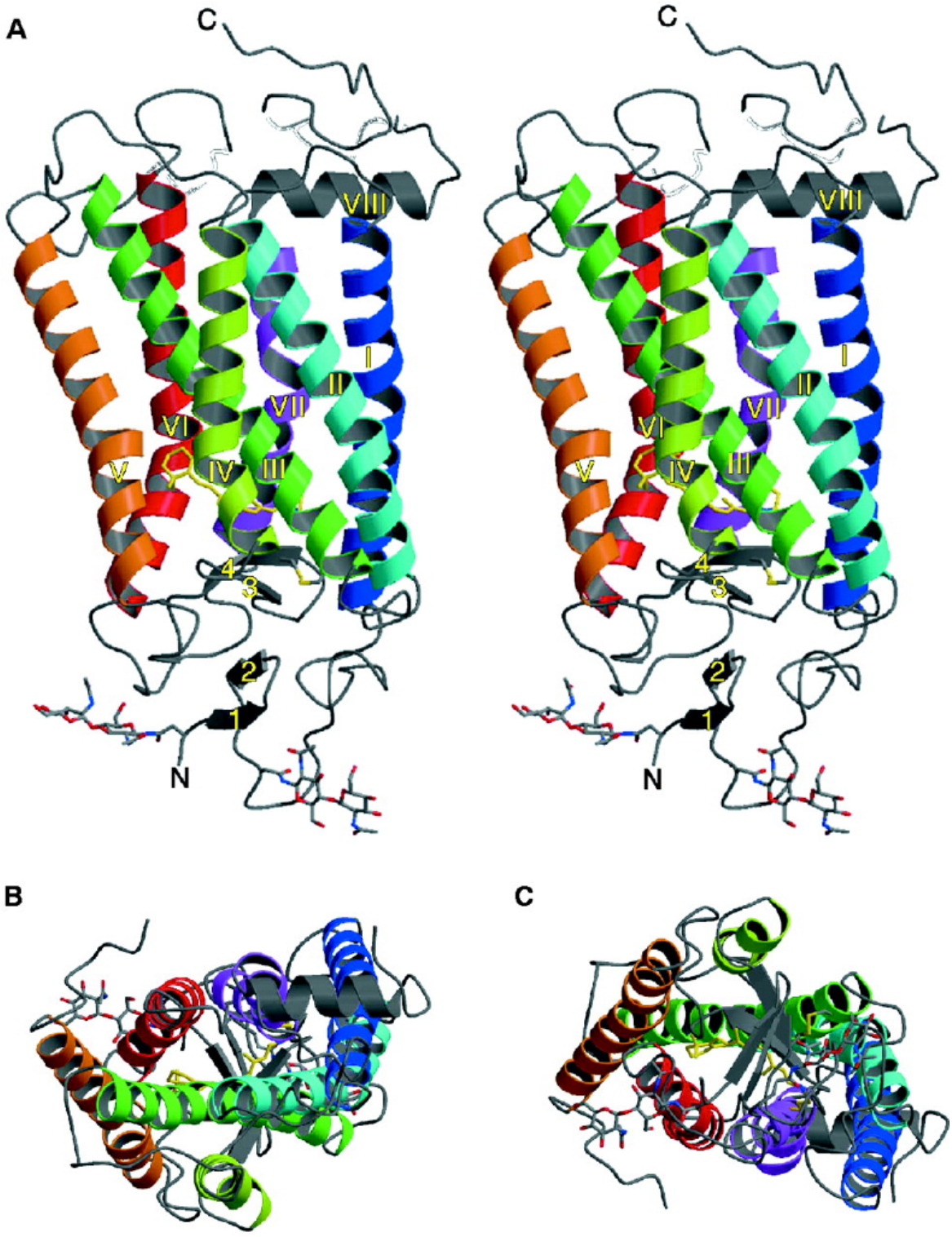


Figure 2.9: Ribbon drawings of rhodopsin of Palczewski et al. [5] (A) Parallel to the plane of the membrane (stereoview). (B) View from the cytoplasmic side and (C) view from the extracellular side orthogonal to the membrane plane.

2.2 CCK-receptors

Cholecystokinin (CCK) is a peptidic prehormone formed by 115 amino acids. It is expressed in neurons throughout the central and peripheral autonomic nervous systems and in intestinal endocrine cells and neurons, where it is released in response to a meal with proteins and lipids. The major forms are the 58- and 8-amino acid C-terminal amidated peptides. The last seven amino acids from the C-terminus are fully responsible for the biological activity. The tyrosine at the seventh position from the C-terminus has to be sulfated for biological activity and it is also essential that the C-terminus is α-amidated [69, 29, 70, 71].
Comparable to CCK, Gastrin also occurs in multiple forms composed by 101 amino acids. The major forms are the 34- and 17-amino acid C-terminal amidated peptides. Although also for Gastrin the α-amidation is essential for biological activity, only the last five amino acids from the C-terminus are necessary. The crucial tyrosine, here located on the sixth position from the C-terminus, has not necessarily to be sulfated (only $\sim$50% are sulfated) [70, 71]. In Figure 2.10 the endogenous ligands for the two type CCK-receptors are depicted and their activity for both receptors are listed.

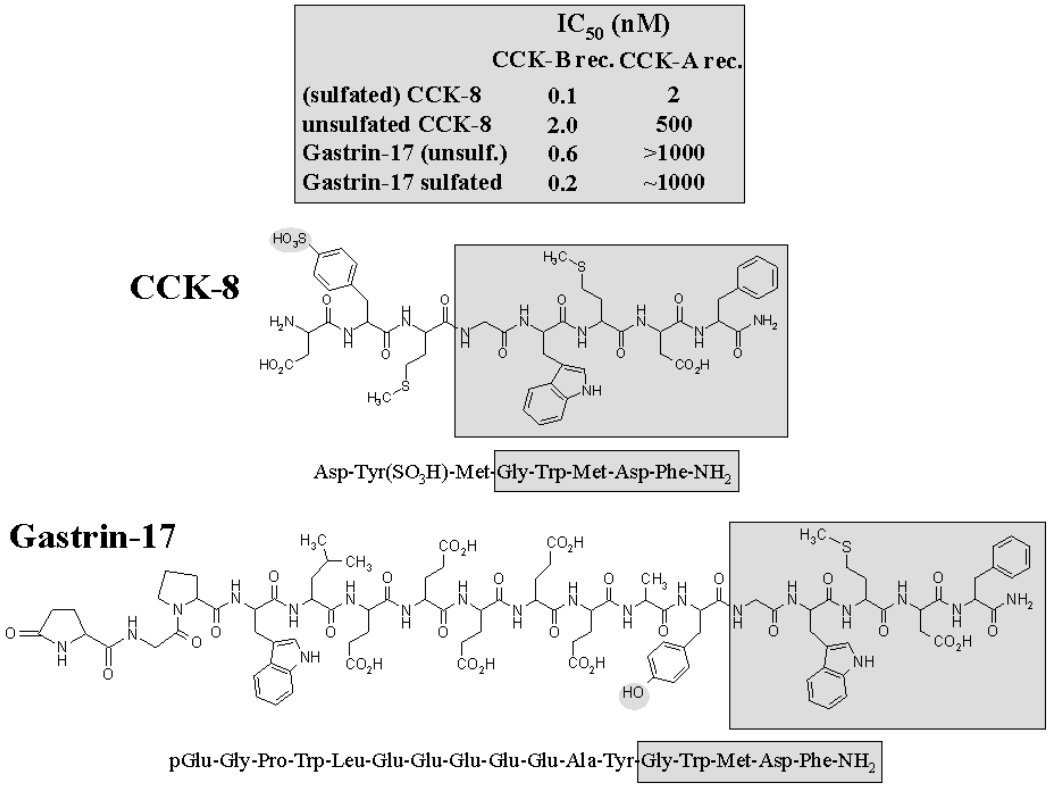


Figure 2.10: Endogenous Ligands of the CCK-A and CCK-B receptors.

The first CCK-receptors were characterised on pancreatic acinar cells and are called CCK type A receptors. In the same year, a second receptor with a different pharmacology was discovered in the brain, the CCK type B receptor. These two receptor types can be pharmacological distinguished on their basis of their affinity to the different CCK and gastrin peptides. The CCK-A receptors are highly selective for sulfated CCK-peptides, whereas the CCK-B receptors have simi-

larly high affinity for both, the sulfated and non-sulfated CCK- and gastrin-peptides [70, 71, and references therein].
CCK is expressed in the central nervous system and the gastrointestinal tract. Endogenous CCK is a primary hormonal regulator of gall bladder contraction in response to a meal and has further great influence on the gastrointestinal physiology as, e.g., pancreatic exocrine secretions of amylases, lipases and proteases, insulin release, and basal acid secretion. At the central nervous system level, CCK is presented as an important neurotransmitter and neuromodulator, involved in the aetiology of anxiety and panic disorders, affecting cognitive functions and release of other neurotransmitters as dopamine [70, 71, and references therein].
Gastrin causes many of the changes mediated by CCK in various tissues and also regulates the release of acid by directly activating CCK-B receptors on parietal cells and on nearby enterochromaffin-like (ECL) cells that release histamine to stimulate acid secretion via H_2-histamine receptors on parietal cells [70, 71, and references therein].

2.3 Molecular Dynamics

Molecular Dynamics (MD) is a simulation technique that allows to investigate the dynamics of molecules. Knowledge of the motions of a complex system of particles, as, e.g., a protein solved in water, provides valuable insights into its functional details. The application of the concept of kinetic energy (velocity) allows the system to overcome energy barriers between nearby local minima and to search considerable areas of the corresponding molecular hypersurface.
From the interaction potentials that operates between the different particles of the system the interaction forces can be computed by calculating the first derivative of the potentials with respect to the positions of the particles. The numerical integration of the Newton's equations of motion yields the new molecular coordinates after the (usually) empirically derived force has been applied for a predefined time period. The leapfrog algorithm is the most common method to accomplish this integration. Iteration of the integration until a specified time period has passed provides the trajectory of the configuration of the system as a function of time. The time step of integration for macromolecules is in the order of a femtosecond (1 fs = 10^{-15}s) to ensure the numerical stability of the integration.
The trajectory can be analysed and different ensemble averages can be determined as, e.g., the average total, potential and kinetic energy, or the average position of an atom. Also radial distribution functions, which give the probability of finding a particle at distance r from another particle, or system properties as, e.g., the temperature or root mean square deviation with reference to the starting structure or atom positions as a function of time can be evaluated. With these statistical mechanics calculation the macroscopic behaviour is analysed by calculating microscopic interactions. Thereby, the assumption is made that the set of configurations is statistically representative for the whole configuration space experienced by the investigated molecular system.
The general process of a molecular dynamic simulation is depicted in the flow chart in Figure 2.11.
The available computer hardware resources (memory and disk space) limit the number of particles of the system and the time regime covered by a MD-simulation. By using periodic boundary conditions (PBC) a simulation can be performed by eliminating unreasonable effects resulting from arbitrary boundary settings. A defined assembly is centred in a periodic box and has iden-

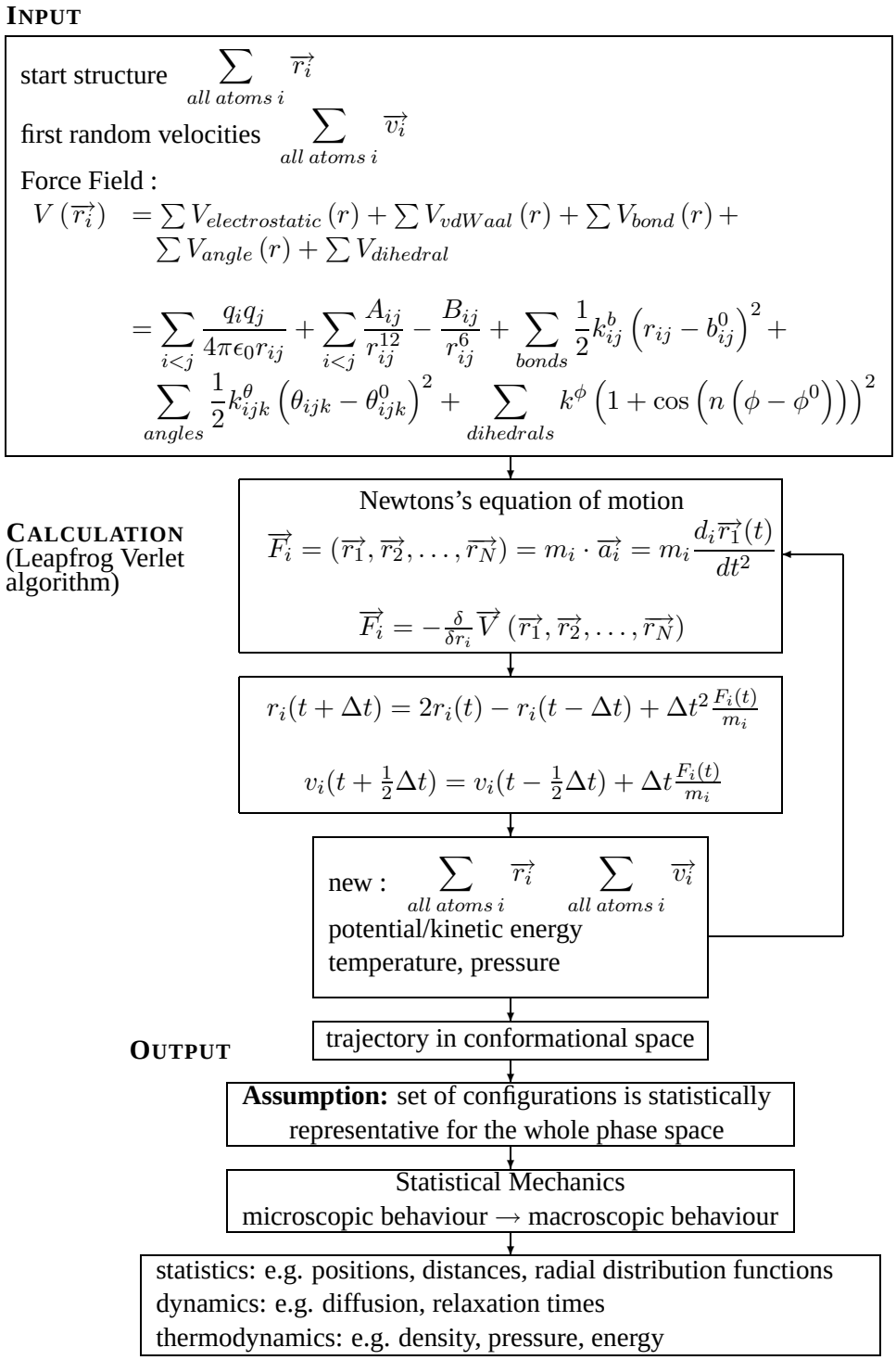


Figure 2.11: A flow-chart of a Molecular Dynamics simulation.

tical neighbouring images on each face of the box. These box images experience the same forces and thus perform the same particle movements as the central box. If a particle moves out of the central periodic box during the simulation, an identical particle enters the box from the opposite face. To avoid the anisotropy of interaction due to the shape of the nearest image box, a spherical cut-off radius is applied. Interactions between pairs of atoms that are further separated than the cutoff radius are set to zero. The cut-off radius should be smaller than one half of one length of the smallest face of the periodic box.
We have applied molecular dynamic simulations for the refinement of structural models which were generated by model-building. Furthermore molecular dynamic simulations are applied to study the flexibility and motion of molecules as, e.g., the lateral diffusion of lipids within a membrane, or the conformational sampling of molecular arrangements of a molecule as a function of the environment (water, non-polar solution etc.).
The methodology, the computational procedures, the simplifications, the assumptions and the approximations of molecular dynamic simulations are reviewed in great detail by van Gunsteren & Berendsen [72] and by Karplus & Petsko [73].

2.4 3Phasic-Box

MD simulations are performed to characterise the dynamics of biologically relevant molecules, in our case the 3D model of the human CCK-B receptor. However, if in such MD simulations an explicit membrane model with explicit lipid molecules and water molecules is used, the size of the system easily exceed the available computational resources. Therefore, a membrane mimetic has been applied in the calculations described in this thesis. This membrane mimetic is described as Lennard-Jones particles corresponding to the physicochemical properties of CCl_4-molecules [74, 75]. They represent a phospholipid bilayer membrane in the liquid-crystalline Lα-phase. This lipid phase is flanked by two water compartments, thus representing different solvent phases of opposing polarity, viscosity, and hydrogen-bond donor/acceptor capabilities. Since the CCl_4-molecules are treated as united atoms, the number of atoms of the simulation cell, in comparison to an explicit membrane model, is reduced by at least one order of magnitude [74, 75]. However, the system should be envisioned as just a crude representation of the complex membrane system of water molecules and phospholipid molecules with their zwitterionic head groups and alkyl chains.
All MD simulations in this thesis were carried out in a rectangular cell with standard three dimensional periodic boundary conditions (PBC) with temperature and pressure bath coupling [72]. After positioning of the modelled protein in the simulation cell, the remaining space in the cell was soaked from previously equilibrated solvent boxes, first with water up to the required height and the rest with CCl_4. The interface plane corresponds to the XZ plane, while the Y-direction is orthogonal to the interface plane of the cell ("the membrane plane"). In case of the receptor simulations, a so-called three-phasic box (Figure 2.12) was generated. In general, the probability distribution of atoms shows an average occupancy along the Y-direction, whereas the time course of the individual Y-coordinates of the atoms illustrates the transition dynamics across the membrane plane. The density profiles of water and CCl_4 can be calculated from the trajectory by counting the number of oxygen and carbon atoms, respectively, in cross-sections at certain heights in the Y-direction and normalised with respect to the idealised atomic densities of water (density 0.0333 number of water-molecules / $Å^3$) and CCl_4 (density 0.0062

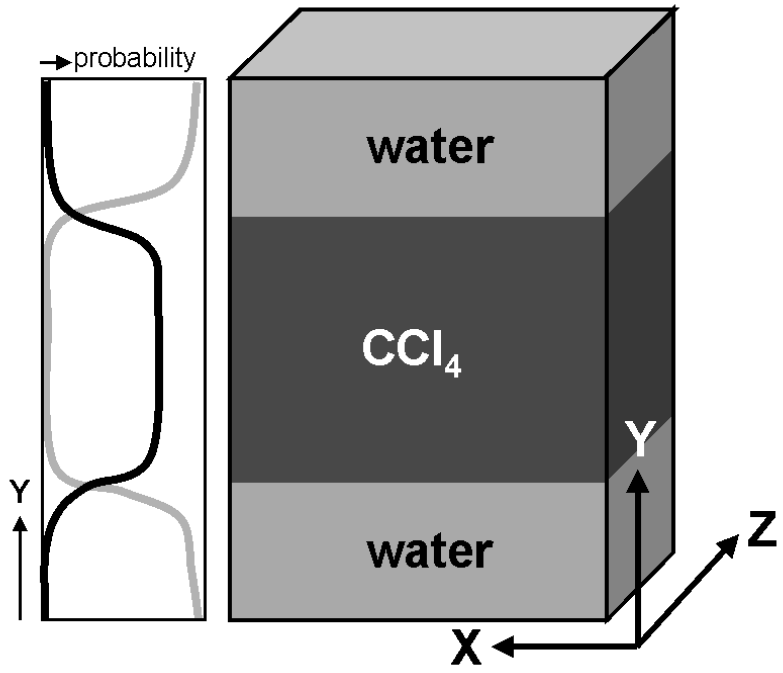


Figure 2.12: The three-phasic box with two water phases and one CCl_4 phase. On the left of the box are the probabilities displayed.

number of CCl_4-molecules / $Å^3$) [74, 75]. These density profiles are overlapping, because of the lower particle density of the CCl_4-molecules in comparison to the particle density of water, and because of a small amount of phase intermixing (Figure 2.12).
Further details about the simulation set-up are given separately within the corresponding sections of this thesis.

Chapter 3

Bioinformatics

Protein Sequence Analysis: “Only a first step towards modelling of G protein-coupled receptors”.

"The Kyte-Doolittle scale is the most widely used in the literature, but its popularity is clearly not sustained by its validity."

- M. Crimi and M. Degli Esposti
  TiBS (**1991**) 16: 119

3.1 Introduction

Extensive biophysical, biochemical and immunological studies have revealed an enormous amount of structural data on GPCRs. These data support that the GPCRs share a common protein topology consisting of a membrane-spanning seven-helix bundle which is believed to accommodate the ligand-binding site, at least for the low-molecular weight compounds. The amino terminus of the membrane protein is located on the extracellular side of the membrane, while the carboxy terminus is situated intracellularly. The transmembrane helices are interconnected by three extra- and three intracellular loops. The intracellular parts of the receptor are involved in binding the G protein, thus giving rise to the name to the receptor family. Despite of all this knowledge, a high-resolution structure of any member of this superfamily was not available when the studies described in this thesis were initiated. In August 2000, researchers found a way to circumvent the difficulties associated with purification and crystallisation of these proteins. Consequently, the high resolution 3D structure of bovine rhodopsin in complex with retinal was solved [5]. However, these results did not serve as experimental structure basis for the sequence analyses described in the following.

At the beginning of this thesis, a low-resolution two-dimensional (2D) map of the electron density of bovine rhodopsin, a GPCR accommodating the visual pigment retinal, provided first experimentally derived structural evidence for the seven transmembrane helix topology for a GPCR [57]. A low-resolution three-dimensional (3D) structure of bovine rhodopsin [63, 64] together with projection structures of other rhodopsins [58, 62, 59] confirmed the arrangement of the seven transmembrane helices. The low-resolution structures also show that, although the membrane topology of bacteriorhodopsin (at that time the only membrane protein with a seven transmembrane helix bundle and a high resolution 3D structure) and the GPCRs are similar, the 3D structures are significantly different. The facts that bacteriorhodopsin is not a GPCR (light driven ion-pump) and displays only a low sequence homology to any of the GPCRs, renders it hard to rationalise that bacteriorhodopsin could be used as template structure for generating 3D models of GPCRs by molecular modelling, even though this has be the usual practise for several years. Only in a few studies the projection maps and low resolution 3D structure of rhodopsin were used, or alternatively, the 3D models were constructed de novo. However, all studies are based on mutually different experimental and theoretical data sets employing different methodologies for receptor reconstruction, thus rendering a comparative judgement of the models impossible.

The first and most critical step in the modelling procedure of every GPCR is the assignment of the transmembrane sequence stretches (TMSS) from multiple protein sequence alignments. Therefore, in this study a comparative evaluation of protein sequence analyses tools was carried out. Several of these tools are tailor-made for the modelling of the seven transmembrane domains (TMDs) of GPCRs. Protein sequence analysis tools such as periodicity analyses of amino acid properties, multiple sequence alignments, distribution analyses of amino acids or directional helix descriptors of, e.g., hydrophobicity or conservation moments were especially developed for GPCR modelling. In this context, we demonstrate that, with respect to the application of different methods, contradictory results can be obtained for the identification of the putative TMSS of GPCRs, as exemplified by a comprehensive protein sequence analysis study based on the most prominent members of peptide-binding GPCRs.

3.2 Methods

3.2.1 Amino Acid Sequences

All amino acid sequences used in this study were retrieved from the protein sequence database SWISSPROT using the computer programs from the University of Wisconsin Genetics Computer Group (GCG [1]). The available sequences of ten different families of peptide-binding GPCRs were retrieved from the database:

	family	# of sequences	# of subtypes
1.	angiotensin II	16	3
2.	bradykinin	5	2
3.	cholecystokinin/gastrin	9	2
4.	endothelin	10	3
5.	interleukin-8	13	2
6.	neuropeptide Y	12	5
7.	somatostatin	17	5
8.	tachykinin	3	
9.	substance P and K	11	2
10.	neuromedin	5	2
	total sequences	101	

For each family, a multiple sequence alignment was generated using the pam250 comparison table [76, 77]. First, the potential transmembrane sequence stretches (TMSS) could be roughly identified from these 10 multiple sequence alignments by the use of the fingerprint of Baldwin [60, 6, 66]. This fingerprint was initially based on 204 sequences of GPCRs [60], including 76 cationic amine receptors, 32 visual pigments, 9 glycoprotein hormone receptors, 66 receptors for peptides and 21 receptors for other small ligands. Baldwin [66] extended this database to 493 sequences leading to an extension of the transmembrane segments by 7 amino acids on both sides to a total length of 40 amino acids. This fingerprint is considered as a relevant indicator for conserved patterns in transmembrane sequences over the entire rhodopsin-like GPCR superfamily. Therefore, is it not surprising that it is found in all 10 multiple sequence alignments. The potential transmembrane sequence stretches (TMSSs) of the 10 different families are further aligned using the fingerprint suggested by Baldwin to generate 7 multiple sequence alignments of 101 sequences, each representing a distinct transmembrane helix.
The lengths for the potential TMSS in these seven multiple sequence alignments were set to 36 amino acids, corresponding to 10 helical turns, which is sufficient to span the hydrophobic core of the membrane. Further, 10 helical turns accounts for a certain margin of error in the fingerprint, i.e., depending on the angle between the helical axis and the membrane plane, only 16 to 18 residues are needed for spanning the hydrophobic core of the membrane.

3.2.2 Amino Acid Property Scales

The seven multiple sequence alignments of 101 sequences, each comprising 36 amino acids, are used to determine the transmembrane domain boundaries (TMDBs) employing different bioinformatic tools and protein modelling techniques. The most common method for the determination of the TMDBs is the hydrophobicity profile of one sequence based on the hydropho-

[1] GCG ® Wisconsin Package, Copyright ©1982-2002 Accelrys Inc., http://www.accelrys.com.

bicity scale of, e.g., Kyte-Doolittle [78]. However, many additional amino acid property scales and other useful approaches exist that address the problem of determination of the TMDBs solely from the primary structure. A comparison of the profiles has been described by several authors [79, 80, 81, 82, 83, 84].
In general, an amino acid property scale can be described as assigning distinct weights to each type of amino acid based on its physiochemical character. For a hydrophobicity scale, this weight represent the tendency of the amino acid to dissolve in water and in nonpolar solvents, respectively. The 20 commonly occurring residues are classified with respect to their side chains as polar, nonpolar, or amphiphilic (partly polar and partly nonpolar). However, even residues with a nonpolar side chain can be reviewed as amphiphilic, because the backbone N-H and C=O groups are themselves highly polar. The method of such a classification of the amino acids can be of theoretical nature, e.g., based on the calculation of the partition coefficients and free energies of transfer, but can also be of more empirical character, e.g., based on the statistical analysis of the observed distributions of the amino acid residues between the water-accessible surface and the buried interior in globular proteins with known 3D structure. The third group of scales, including the scale of Kyte-Doolittle [78], is a combination of different scales. As a matter of fact, these scales of different origin can disagree substantially, depending on the calculated method, the criteria used and the investigators involved. These differences are not minor disagreements in the relative ordering of hydrophobicity of the amino acids, but they extend to the classification whether an amino acid has to be considered as hydrophilic or hydrophobic. The choice of the amino acid property scale will have severe influence on the results for the determination of the TMDBs.
Since only a small number of high resolution 3D structures of membrane proteins is known (bacteriorhodopsin, photosynthetic reaction centre, porin, potassium channel and in some extent: the light-harvesting complex, the acetyl choline receptor and rhodopsin) the accuracy of the different amino acid property scales used in the different determination methods of the TMDBs cannot be evaluated by using a large test-database. Consequently, it can hardly be assessed whether the prediction error of the TMDBs depends on the choice of the amino acid property scale, the selected determination method or whether it is simply a consequence of the nature of the proteins of this small set itself. Therefore, an evaluation of different bioinformatic tools that use these amino acid property scales for the determination of the TMDBs of GPCRs, has to complemented by as much experimental information as possible.

3.2.3 Amino Acid Property Profiles

The amino acid property scales used in this study are listed in Table 3.1. The chosen amino acid property scales are used in different bioinformatic tools, e.g., in the calculation of a property profile of an amino acid sequence. With the selected amino acid property scale, an average property is calculated for a sequence window. The plot of this average property of the sequence window against the sequence number of the central amino acid results in the property profile. The choice of the length of the sequence window is related to the thickness of the membrane bilayer. Depending on the angle between the main helix axis and the membrane plane, the transmembrane helix needs to be composed of minimally 16 to 18 residues in order to span the hydrophobic core of a general lipid bilayer at the level of the carbon chains of the phosphatidyl lipids. However, this thickness of the membrane bilayer can vary considerable, therefore in many studies the optimal window was set to 19. Also smaller windows (for example 7) have

Method	Scale	Refs.
variability/conservation of amino acids in high homologue sequence groups	- fingerprint analysis of Baldwin	[60, 6, 66]
	- conserved polar residues of Zhang and Weinstein	[85]
amino acid property scales for property profiles	- hydrophilicity of Kyte-Doolittle	[78]
	- hydrophilicity of Wolfenden	[86]
	- hydrophilicity of Eisenberg	[87]
	- hydropathy by Hopp and Woods	[88]
	- hydrophobicity by Abraham	[89]
	- hydrophobicity of Roseman	[90]
	- hydrophobicity of Sweet	[91]
	- polarity index of Zimmerman	[92]
	- bulkiness index of Zimmerman	[92]
	- mutability index of Dayhoff	[76]
	- burriness index of Janin	[93]
	- helical index of Chou/Fasman	[94]
amino acid occurrence within the transmembrane regions	- positive inside rule by von Heijne	[95]
	- proline residues in transmembrane helices	[96, 97, 98]
	- aromatic residues in membrane boundary regions	[98]
α-helical periodicity of amino acid properties	- helical wheel representation	[99]
	- sequence property moments	[100, 101, 87]
	- helical net representation	[102]
	- Fourier Transform analysis	[103]

Table 3.1: Amino acid property scales and approaches for the determination of the Transmembrane Domain Boundaries.

been used, since they give a better resolution to separate TMDs and a better prediction of the TMD-termini. In this study, we employed several sequence window sizes; 3, 4, 5, 6, 7, 9, 11, 17 and 19, in the calculation of property profiles.
Given the fact that a mutation of a single amino acid residue can significantly influence the property profile of a single sequence, we have not calculated the property profiles of single sequences, but calculated property profiles for the different multiple alignments of the seven TMSS. These property profiles are thus average profiles over multiple sequences and have the advantage to compensate for small, irrelevant deviations of single amino acid sequences. For each TMSS, each property scale and each sequence window 11 different property profiles were calculated; 10 of the multiple alignments of the 10 different families and one of the multiple alignment of all 101 sequences. This results in 11 (alignments) * 12 (scales) * 9 (windows) = 1188 profiles per transmembrane sequence, from which the TMDBs were determined. To reliably identify a TMDB, the property profile must exhibit a clear break at this point. A TMDB is clearly identified whenever a distinct discontinuity occurs in a property profile.

3.2.4 Amino Acid Distribution Profiles

Furthermore, also profiles for the distribution of the 20 common amino acids occurring in the different multiple alignments of the TMSSs were determined. This parallels the study of Landolt-Marticorena [98] on the distribution of amino acids in the transmembrane segments and flanking regions of human type I single span plasma membrane protein. In agreement with the positive inside rule of von Heijne [95, 104], Landolt-Marticorena [98] reported that the amino acids are non-randomly distributed over the transmembrane sequence and their occurrence depends on the topology of the TMD. In agreement with these findings, it can be expected, that these non-random-distribution of amino acids will also be found in the multiple alignments of the TMSSs of GPCRs. By depicting the distribution profiles as the observed percent occurrence of each amino acid at a position in the TMSS, these profiles can also be a tool for the determination of the TMDBs. These distribution profiles were determined for the seven multiple alignments of the TMSSs with all the 101 sequences. For comparison and validation, distribution profiles were also constructed for seven multiple alignments of the TMSSs of 50 sequences of the family of opsine receptors, another important member of the superfamily of GPCRs.

3.2.5 Helical Property Moments

The different amino acid property scales can be further used in the determination of the so-called property moments of the TMDs [100]. A property moment is a quantitative description of the asymmetry of the used property over a particular molecular structure, for example, a sequence of amino acid residues with a defined secundary structure. Using a scale of atomic charges over a certain molecular conformation will determine the electric dipole moment as property moment. For the calculation, an ideal α helical conformation is assumed for the TMSSs. Already in 1967, Schiffer and Edmundson represented an ideal α helix as a two-dimensional helical wheel [99]. The property moment for an ideal α helix of amino acids can be calculated by representing the hydrophobicity of an amino acid as a vector directed radial from the projected helix centre to the projected idealised Cα-atom position, with the length of the vector being the property value of this amino acid. The sequence helical property moment is then given by:

$$\mu_{HP} = \left\{ \left[\sum_n P_n \sin(\delta n) \right]^2 + \left[\sum_n P_n \cos(\delta n) \right]^2 \right\}^{\frac{1}{2}} \tag{3.1}$$

The sum, μ_{HP} accounts for all amino acid residues of the sequence segment, P_n is the property of the n^{th} residue, and δ is the angle (for an ideal α helix 100°), measured in radians, between consecutive Cα-atoms.

In this study, the helical property moments have been calculated with different property scales for the TMDs of the 11 different multiple alignments of each of the seven TMSSs. Again, the consensus approach will rule out the influence of an irrelevant deviation by a single amino acid mutation in a sequence. In contrast to the polarity moments determined by Zhang and Weinstein [85], these property moments are based on the property character of the total TMD and not on single, highly conserved residues.

For comparison and better validation, also the sequence property moments were calculated for the multiple alignments of the seven TMDs of 50 opsine receptors and additionally for the multiple alignments of the seven TMDs of 623 sequences of rhodopsin-like GPCRs. These

seven multiple alignments of 623 sequences were subtracted from the alignments given on the web-pages of the GPCRDB-server at EMBL-Heidelberg [2].
By molecular modelling, a 3D model of a GPCR can be build using the detected TMDs. We used the Cα trace model of Baldwin [66] to transfer the sequentially derived information into a rough 3D model. From that, structural property moments can be calculated for the transmembrane helices. The atomic co-ordinates of the amino acids are used and not just an ideal projection of them on a helical wheel representation. The structural helical property moment is given by:

$$\mu_{SP} = \sum_n P_n v_n \tag{3.2}$$

The sum, μ_{HP}, accounts for all amino acid residues of the sequence segment, P_n is the value of the property of the n^{th} residue, and v_n is the unit vector pointing from the Cα-atom of the n^{th} residue to its Cβ-atom. This moment has the advantage that it refers to a realistic α helix structure with bends and kinks and not an ideal α helix.
In order to get detailed insights into the comparability of sequence helical property moments and structural helical property moments, we determined these property moments for two proteins with known high resolution 3D structure comprising a helical topology, notably bacteriorhodopsin (PDB: 2brd) and the four helix bundle protein Felix (PDB: 1FLX and 3FLX). The later is a designed globular protein, which should have an ideal distribution of hydrophobic amino acid residues in the interior and hydrophilic amino acid residues on the surface, thus reflecting an opposite distribution of properties when compared to the transmembrane GPCR proteins.

3.2.6 Perscan and Other Tools

Apart from the application of, e.g., helical net representations [102], specially developed neural networks by B. Rost [105, 106] [3], the computation of helical moments is implemented in Perscan [107, 67, 108]. Both are claimed to be useful in predicting TMDs. The reliability of the later approach for the purpose of GPCR modelling was also tested in this study.

3.3 Results

3.3.1 Transmembrane Sequence Stretches

The TMDs of membrane proteins are usually determined using a hydrophobicity plot of the individual sequence of the membrane protein, generally based on the hydrophobicity scale of Kyte-Doolittle [78]. The plot of the average hydrophobicity against the sequence number identifies regions of the amino acid sequence with overall hydrophobic character such as expected of a TMD. In general, a hydrophobic region is identified as a TMD if it spans at least 18 amino acids. In this study, this procedure was carried out for all nine sequences of the CCK-A & B receptors, not only based on the hydrophobicity scale of Kyte-Doolittle, but also applying

[2] now moved to the Center for Molecular and Biomolecular Informatics (CMBI) in the Netherlands: http://www.gpcr.org/7tm
[3] URL: http://www.embl-heidelberg.de/predictprotein

other amino acid property scales. In Figure 3.1, an example of the profile of the sequence of the human CCK-B-receptor with the polarity scale of Zimmerman and the hydrophobicity scale Kyte-Doolittle with different window sizes is shown. It is obvious, that the TMDs cannot be determined straightforward from this plot, also because the result is strongly influenced by three factors: notably the applied scale, the window size and the threshold value, that determines a sequence as hydrophobic or hydrophilic. Furthermore, only a few mutations within the sequence significantly influences the property profile plot and therefore the determination of the TMDs. Consequently, this method is considered as not practicable for the exact determination of the TMDBs, which is a prerequisite for the molecular modelling of a GPCR.

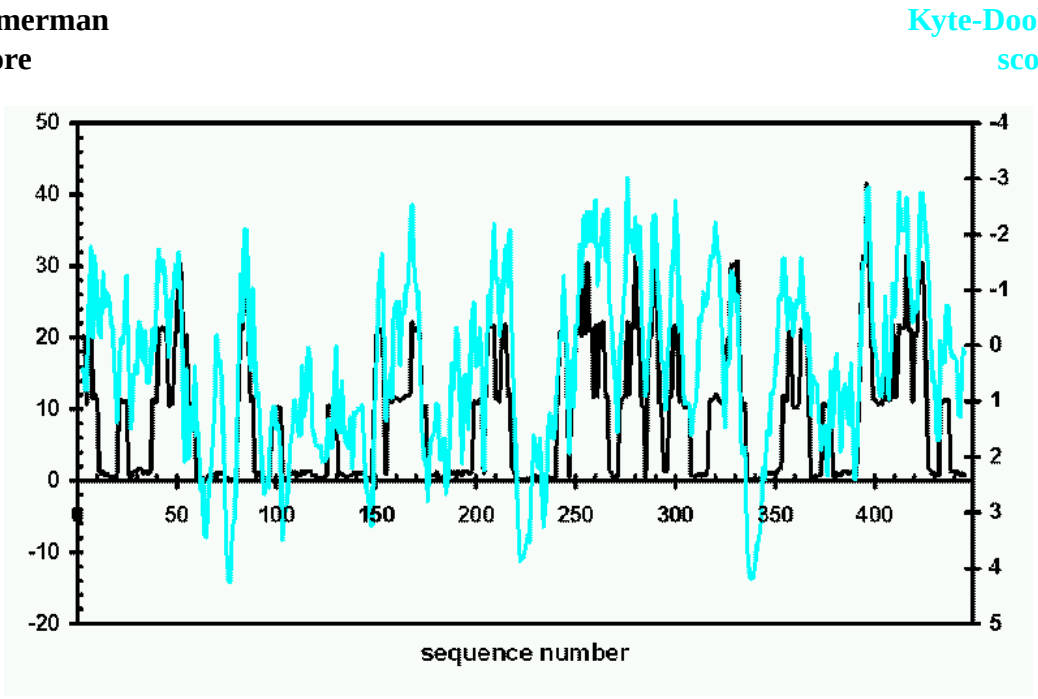
The separate multiple alignments of the 10 different peptide binding GPCR families were generated using the computer programs from the University of Wisconsin Genetics Computer Group (GCG). The resulting 10 multiple alignments exhibit high similarity in a single family, especially in the regions were the TMDs are expected. Supported by the fingerprint of Baldwin [66], the seven potential TMSSs could be determined in all sequences. Strongly conserved amino

Helix II		Baldwin Fingerprint	Baldwin Lines	cckr_human	gast_human	et1r_human	etbr_human	ny1r_human	ny2r_human	ny4r_human	ny5r_human	ny6r_mouse
1	-4	#	5	N	S	N	N	Q	F	Q	K	Q
2	-3	#	10	K	R	K	K	K	K	K	R	R
3	-2	#	15	R	R	C	C	E	S	E	N	E
4	-1	#	8	M	L	M	M	M	M	K	Q	A
5	0	#	3	R	R	R	R	R	R	A	K	Q
6	1	#	5	T	T	N	N	N	T	N	T	N
7	2	+	2	V	V	G	G	V	V	V	T	V
8	3	#	1	T	T	P	P	T	T	T	V	T
9	4	#N(70)	2	N	N	N	N	N	N	N	N	N
10	5	+	4	I	A	A	I	I	F	L	F	I
11	6		0	F	F	L	L	L	F	L	L	L
12	7	I/L(77)	0	L	L	I	I	I	I	I	I	I
13	8	+	5	L	L	A	A	V	A	A	G	A
14	9	#N\S(88)	1	**S**	**S**	**S**	**S**	**N**	**N**	**N**	**N**	**N**
15	10	L(96)	0	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**
16	11	A(86)	3	**A**	**A**	**A**	**A**	**S**	**A**	**A**	**A**	**S**
17	12		16	V	V	L	L	F	V	F	F	L
18	13	A/S(85)	3	**S**	**S**	G	G	**S**	**A**	**S**	**S**	**S**
19	14	#D(94)	0	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**
20	15	+L/F(80)	3	L	L	L	L	L	L	F	I	I
21	16	L	2	M	L	I	L	L	L	L	L	L
22	17	+	3	L	L	Y	H	V	V	M	V	V
23	18	+	0	C	A	V	I	A	N	C	V	C
24	19		16	L	V	V	V	I	T	L	L	V
25	20	+	2	F	A	I	I	M	L	L	F	M
26	21	#	1	C	C	D	D	C	C	C	C	C
27	22		3	M	M	L	I	L	L	Q	S	I
28	23	+	4	P	P	P	P	P	P	P	P	P
29	24	+	4	F	F	I	I	F	F	L	F	F
30	25	+	6	N	T	N	N	T	T	T	T	T
31	26	+	8	L	L	L	V	V	F	L	A	L
32	27	+	11	I	L	F	Y	V	T	V	T	I
33	28	#	8	P	P	K	K	Y	Y	Y	S	Y
34	29	#	8	N	N	L	L	T	T	T	V	T
35	30	+	12	L	L	L	L	L	L	I	L	L
36	31	#	6	L	M	A	A	M	M	M	L	M

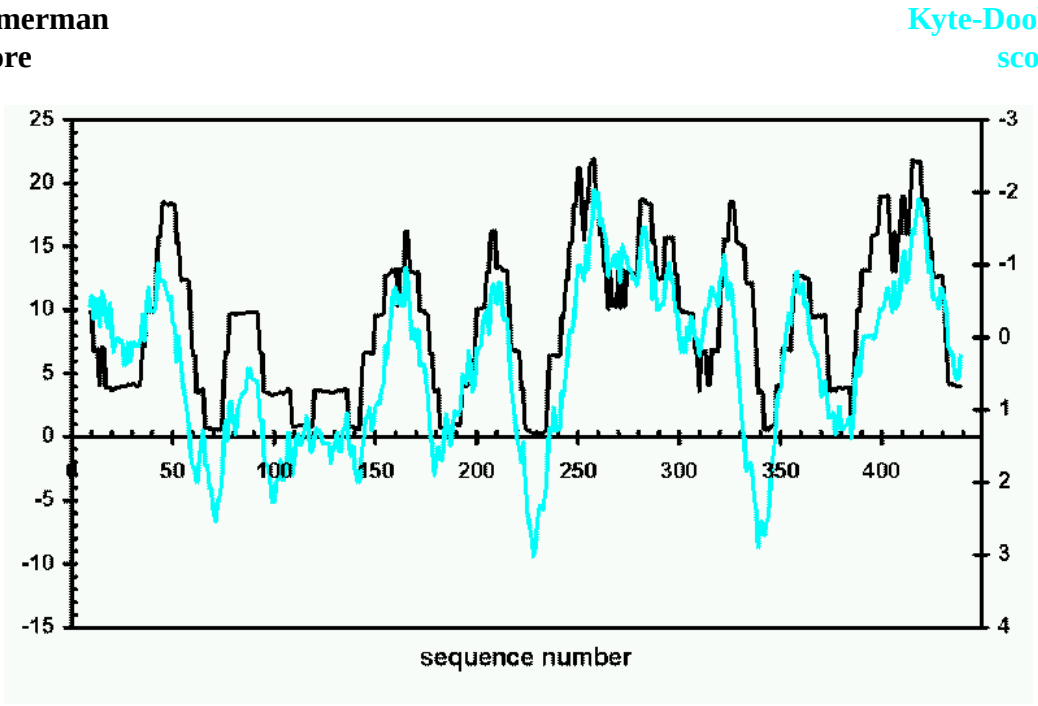
Helix V		Baldwin Fingerprint	Baldwin Lines	cckr_human	gast_human	et1r_human	etbr_human	ny1r_human	ny2r_human	ny4r_human	ny5r_human	ny6r_mouse
	32	#										
36	31	#	16	Q	L	N	R	R	N	R	C	K
35	30	#	11	Y	Y	L	L	K	K	Q	S	R
34	29		0	L	L	M	M	L	L	L	I	L
33	28	#	9	E	E	E	E	R	K	R	S	C
32	27	#	10	L	R	C	C	I	S	R	R	L
31	26	+	2	S	S	T	T	Y	W	Y	C	V
30	25	I/V(72)	1	I	I	M	M	I	I	I	V	I
29	24	#	10	L	L	L	L	K	R	R	S	K
28	23	+	13	G	G	T	T	F	T	A	T	L
27	22	+Y(91)	0	Y	Y	Y	Y	Y	Y	Y	H	Y
26	21		2	A	A	F	F	C	S	C	S	C
25	20	+	14	V	V	I	F	I	F	V	V	I
24	19		11	M	A	A	A	F	S	L	T	L
23	18	M/I(79)	1	M	M	T	T	I	I	I	L	I
22	17		16	V	V	C	I	F	I	F	C	F
21	16		16	I	V	V	A	C	G	G	V	G
20	15	L	5	G	G	L	L	L	L	L	L	L
19	14	P(94)	0	P	P	P	P	P	P	P	P	P
18	13		16	I	I	M	L	G	L	L	L	V
17	12		7	L	F	C	C	F	V	C	I	F
16	11	F(70)	1	F	F	F	F	Y	Y	Y	Y	Y
15	10	#	6	L	L	Y	Y	Q	L	Q	Q	Q
14	9		16	I	L	F	F	L	I	F	V	L
13	8		16	L	L	G	S	V	L	L	L	M
12	7	+	2	L	L	F	F	L	L	L	L	F
11	6	#	11	L	L	L	L	L	S	L	L	L
10	5	+	6	F	L	W	W	L	S	F	S	S
9	4		8	T	V	W	W	T	L	T	I	T
8	3	#	8	H	S	D	D	T	S	T	T	S
7	2	+	4	W	W	K	K	Y	Y	Y	F	F
6	1	+	10	S	T	V	A	S	V	I	A	L
5	0	#	11	Q	Q	D	T	L	T	T	I	L
4	-1	#	9	Q	R	Q	K	R	G	R	R	Q
3	-2	#	6	M	V	Y	Y	H	Y	H	Y	N
2	-3	#	9	V	R	F	F	S	I	H	S	L
1	-4	#	12	D	A	E	Q	D	S	A	D	K

Helix VII		Baldwin Fingerprint	Baldwin Lines	cckr_human	gast_human	et1r_human	etbr_human	ny1r_human	ny2r_human	ny4r_human	ny5r_human	ny6r_mouse
36	31	#	7	G	A	C	C	D	A	E	D	D
35	30	#	16	L	Q	N	N	R	K	K	A	K
34	29	#R/K(79)	5	R	R	K	K	Q	R	K	K	Q
33	28	+F/Y(77)	5	F	F	F	F	Y	F	I	F	
32	27	#	7	R	R	K	R	N	N	G	N	
31	26	#	7	K	R	K	K	K	S	T	N	K
30	25	#	1	N	H	S	S	N	N	N	N	N
29	24	+	1	M	M	V	V	L	M	L	L	L
28	23	+	7	F	F	F	L	F	W	F	F	F
27	22	+	1	C	C	Y	Y	G	G	G	G	G
26	21	Y(95)	0	Y	Y	L	L	Y	Y	Y	Y	Y
25	20		1	I	V	A	A	F	L	I	L	F
24	19		9	I	L	I	I	I	L	F	I	L
23	18	P(98)	0	P	P	P	P	P	P	P	P	P
22	17	#N/D(99)	1	N	N	N	N	N	N	N	N	N
21	16	+	3	V	V	I	I	V	A	V	L	I
20	15		4	C	C	C	C	C	F	C	C	C
19	14	+S/C(76)	2	S	A	S	S	T	T	T	C	T
18	13	#N\S(80)	3	S	S	N	N	S	S	S	S	S
17	12		8	T	A	M	L	I	C	A	M	V
16	11	#	1	Y	Y	T	S	M	M	M	M	M
15	10		7	S	S	A	A	A	A	A	G	A
14	9	L/F(83)	5	L	L	L	M	T	I	L	L	I
13	8	+	2	L	L	N	N	L	I	L	L	L
12	7	#	2	L	H	I	I	H	H	H	H	H
11	6	+	7	I	I	G	G	C	F	C	C	C
10	5	+	9	F	F	I	I	L	V	V	I	V
9	4	#	7	S	S	Y	Y	L	T	L	C	V
8	3	#	4	I	I	D	D	F	F	F	Y	F
7	2	+	13	P	P	M	L	L	I	I	V	V
6	1	#	16	T	A	L	V	L	L	L	L	L
5	0	#	16	G	G	L	L	N	K	N	K	D
4	-1	#	12	S	S	L	L	H	Y	G	F	H
3	-2	#	16	L	L	F	F	N	E	H	H	H
2	-3	#	16	R	A	S	S	C	K	C	R	C
1	-4	#	16	R	R	L	L	T	L	I	N	S

Table 3.2: The fingerprint of Baldwin for the Transmembrane Domains 2, 5 and 7 of the rhodopsin-like family of G protein-coupled receptors is given along with the potential Transmembrane Sequence Stretches of the human CCK-A and B receptors, the human neuropeptide-Y receptors and the endothelin receptors. Yellow-amino acids = inside the receptor, pale-yellow-amino acids = on border of protein, red-numbers = centre amino acids, blue line = end of transmembrane sequence, red line = end of core sequence, and magenta line = possible end of core sequence.



A



B

Figure 3.1: Property profiles of the single amino acid sequence of the human CCK-B receptor with the polarity scale of Zimmerman (black) and the hydrophobicity scale of Kyte-Doolittle (blue, score reversed plotted) in (A) with window size 5 and in (B) with window size 17.

acids, indicated by Baldwin with a percentage of conservation above 70%, were almost always identified in all 101 sequences. In each TMSS, amino acids could be used as anchor points for the aligning of the TMSSs of all 101 receptors for each TMD. In Table 3.2 the fingerprint of Baldwin [66] is given along with the potential TMSSs of the human CCK-A and B receptors, the human neuropeptide-Y receptors, as well as the endothelin receptors.

3.3.2 Amino Acid Property Profiles

For each of the seven TMDs, 11 multiple alignments with a length of 36 amino acids (10 from 10 different receptor families and one for all the 101 receptors) were generated. Property profiles were computed for all 77 alignments with 12 different scales and 9 different window sizes, resulting in total (77 x 12 x 9 =) 8316 profiles.
In order to circumvent the problem of an unambiguously assignment of a meaningful threshold value for, e.g., hydrophobicity, we have normalised all profiles, also to ensure comparability of profiles generated with different scales. For better comparison we have also reversed the score of the profiles for hydrophobicity, thus regions of high hydrophobicity display similar scores when compared to regions of low polarity.
In general, the property profiles for the larger window sizes (> 7) are not useful for identification of the TMDBs, they can only be used for the identification of a TMD in a sequence. In Figure 3.2 the normalised property profiles based on the polarity scale of Zimmerman and the hydrophobicity scale of Kyte-Doolittle for the TMSSs 2, 3 and 4 are depicted for different window sizes (3, 7, and 17 respectively). The profiles with window size 17 are generally flatter and fail to display a high resolution, which is required to determine the TMDBs. On the other hand, the profiles with the smaller window size of 3 are more noisy and have occasionally discontinuity in their plots, which renders determination of the TMDBs questionable.
In the following, thorough sequence analysis is based on the profiles derived from the multiple alignment of all 101 sequences, thus revealing a more general picture of sequence characteristics, less influenced by mutations within distinct sequences or less populated families.
From the normalised profiles shown in Figure 3.3, it turns out that the bulkiness scale of Zimmerman, the helical parameter of Chou-Fasman, the mutability scale of Dayhoff, and the scale of buried residues of Janin do not produce distinct regions of approximately 18 amino acids length with a significant decreased or increased value as diagnostic indication for the existence of a TMD. The profiles vary over the entire TMSS and no consistency among the different TMDs can be obtained. Consequently, the scales of Janin, Chou-Fasman, Dayhoff and the bulkiness scale of Zimmerman cannot be used for the determination of the TMDBs, while the hydrophobicity scale of Kyte-Doolittle, Eisenberg, and the polarity scale of Zimmerman produce more promising results.
No significant differences between the profiles generated with window sizes 3, 4 and 5, and between 6, 7 and 9 could be determined. After the optimisation process discussed above, 8 scales (the hydrophobicity scales: Kyte-Doolittle, Abraham, Eisenberg, Roseman, Sweet and Wolfenenden; the hydropathy scale of Hopp-Woods and the polarity scale of Zimmerman) and 4 window sizes (3, 7, 11 and 17) were chosen for further analysis.
The profiles of the hydrophobicity scales are not always identical, but they have in general the same pattern (see e.g. TMSS 7 in Figure 3.3B). Figure 3.2 and 3.3 clearly proves that the hydrophobicity scales do not allow a straight identification of potential borders of the TMDs, they display more than one discontinuity in the profile and they not necessarily encompass

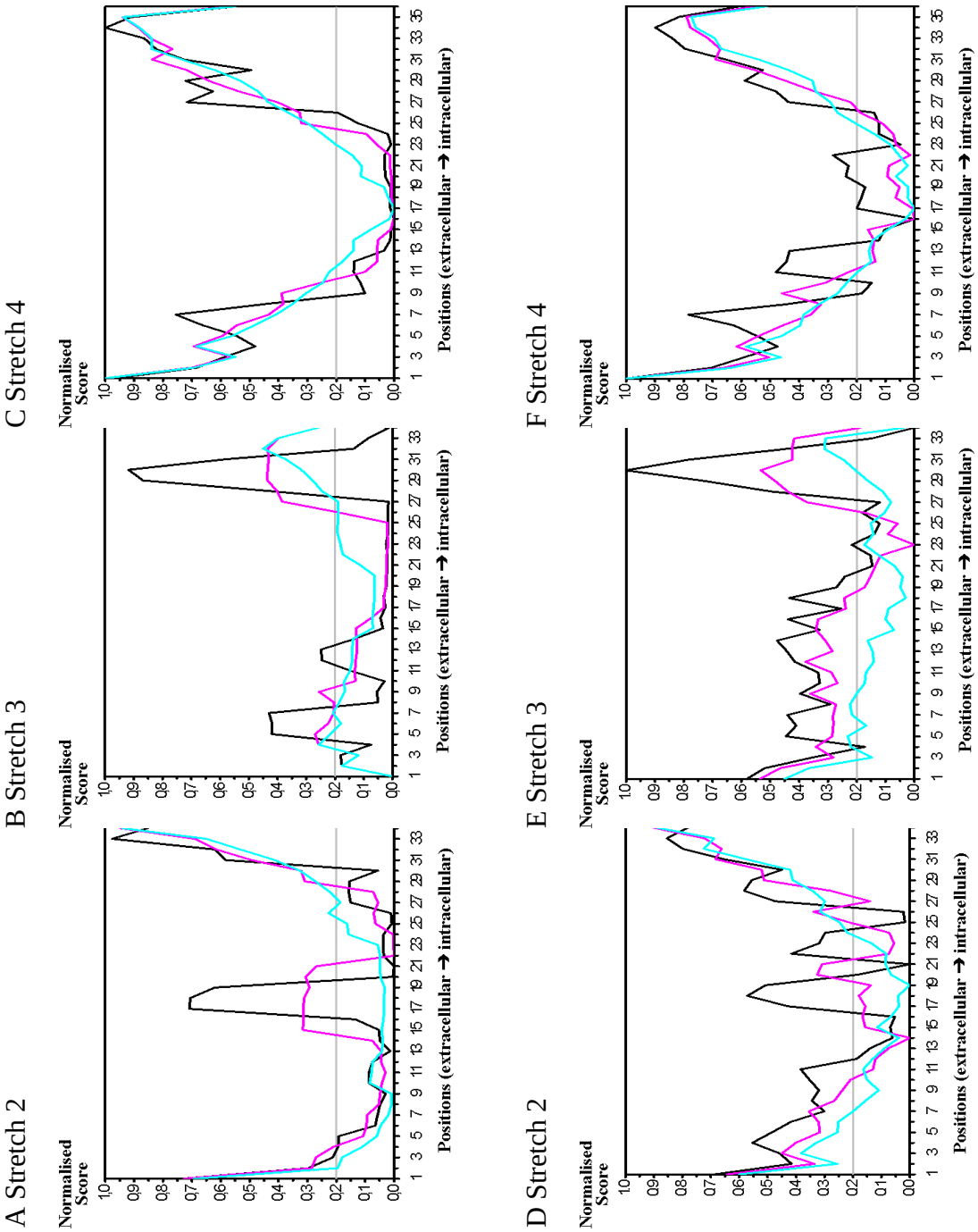


Figure 3.2: The property profiles of the Transmembrane Sequence Stretches 2, 3 and 4 with a window size of 3 (black), 7 (magenta) and 17 (blue) and in the first row with the polarity scale of Zimmerman (A, B and C) and in the second row with the hydrophobicity scale of Kyte-Doolittle (D, E and F).

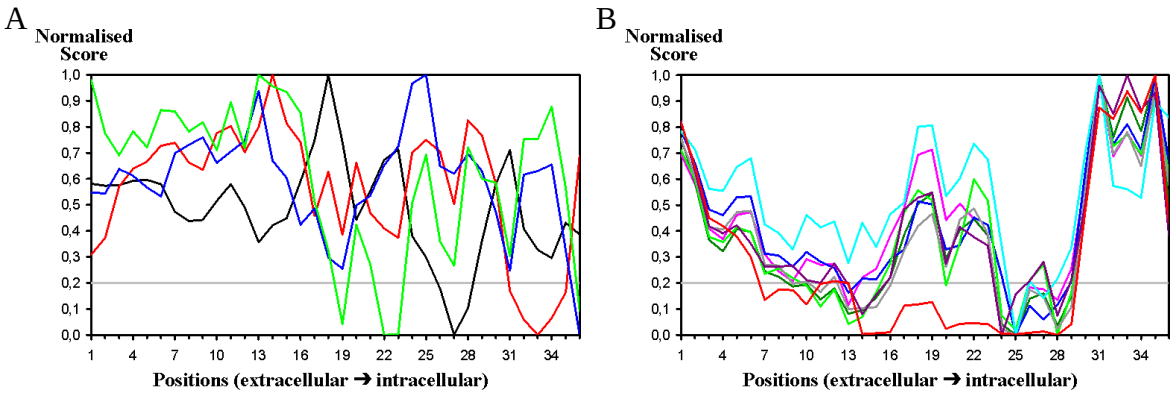


Figure 3.3: The property profiles with a window size of 3 of Transmembrane Sequence Stretch 2 with in A : the scales of Zimmerman for bulkiness (blue), the helical parameter scale of Chou-Fasman (green), the mutability scale of Dayhoff (black) and the scale of buried residues of Janin (red). In B the polarity scale of Zimmerman (red), the hydrophobicity scales of Wolfenden (grey), Eisenberg (black), Abraham (pink), Kyte-Doolittle (light green), Roseman (dark green) and Sweet (light blue), and the hydropathy scale of Hopp-Woods (dark blue).

18 amino acids with a clear decreased normalised value. This phenomenon occurs in a much lesser extent in the profiles of the polarity scale of Zimmerman. Therefore, the profiles with the polarity scale of Zimmerman are used as basis for the determination of the TMDBs. Border residues are identified as flanking stretch of approximately 18 amino acids with a normalised value ≤ 0.2 in the profile of Zimmerman, determined with a window size of 3. It turns out that in TMSSs 3, 6 and 7, these stretches of amino acids with a value smaller than 0.2 are interrupted at the extracellular side by a small region of 2 amino acids. However, these regions are included in the determined TMD since they are flanked by regions of at least 2 amino acids with a normalised value significantly smaller than 0.2. The determined TMDBs are listed in Table 3.3A according to the numbered positions in the graphs (Figures 3.1-3.3).
Table 3.3B gives also the centres of the TMDs which were determined as the average of the first and the last residues of a region with a normalised value smaller than 0.2 for at least seven of the eight property profiles, generated with a window size 17. For reasons of comparison, the corresponding parameters by Baldwin [60, 6, 66] are also given.

3.3.3 Amino Acid Occurrence Profiles

The occurrence profiles for different amino acids over the TMSSs of both, the family of peptide binding GPCRs, and the family of opsine receptors, show that the distribution of key amino acids is clearly not random, thus depending on the transmembrane protein topology. These results compare well with the findings of von Heijne [95, 109] and Landolt-Marticorena [98].
From the occurrence profiles for arginine and lysine a preference of these positively charged amino acids for TMD-flanking regions on the intracellular side of the TMSSs can be deduced. However, they also occur in a lesser extent at the extracellular side of the TMSSs, as depicted in Figure 3.4A. The occurrence profiles of negatively charged residues are a mirror image of that of positively charged residues. They occur on both sides of the TMDs, with in general clear preference for the extracellular side (normal pattern for peptide binding GPCRs Figure 3.4C, deviating pattern for opsine receptors Figure 3.4D).

A

	TMDB determined by Baldwin				TMDB determined with window size 3						
	amino acid numbers of TMDB in TMSSs. extrac.	intrac.	length of stretch	position of TMSS in graph †	determined with only prop. prof.		determined with prop. prof. + aa distr.		length of stretch	difference to TMDB determined by Baldwin	
Helix 1	10	27	18	10	27	7	30	9	27	19	1	0
Helix 2	27	10	18	10	27	4	30	10	27	18	0	0
Helix 3	10	27	18	10	27	8	27	11*	27	17	-1	0
Helix 4	26*	11	16	11	26	9	26	10*	26*	17	1	0
Helix 5	11*	28	18	11	28	10	30	12	28	17	-1	0
Helix 6	27	11	17	10	26	8	27	10	27	18	0	1
Helix 7	12*	27	16	12	27	7	29	12*	28*	17	0	1

* = some difficulties with the determination of these transmembrane boundaries, see text.

† = the positions of the odd transmembrane sequence stretches are reversed numbered in the graphs.

B

	Middle Points of the determined Transmembrane Domains						
	middle point of the TMDBs determined by Baldwin	middle point of the TMDBs determined with only prop. prof. and window size 3		middle point of the TMDBs determined with prop. prof. + aa distr. and window size 3		middle point of the TMDBs determined with only prop. prof. and window size 17	
		Middle point	difference to Baldwin	Middle point	difference to Baldwin	Middle point	difference to Baldwin
Helix 1	18,5	18,5	0,0	18,0	-0,5	18,0	-0,5
Helix 2	18,5	17,0	-1,5	18,5	0,0	17,0	-1,5
Helix 3	18,5	17,5	-1,0	19,0	0,5	18,0	-0,5
Helix 4	18,5	17,5	-1,0	18,0	-0,5	17,5	-1,0
Helix 5	19,5	20,0	0,5	20,0	0,5	18,5	-1,0
Helix 6	18,0	17,5	-0,5	18,5	0,5	19,0	1,0
Helix 7	19,5	18,0	-1,5	20,0	0,5	19,0	-0,5

Table 3.3: (A) The Transmembrane Domain Boundaries elucidated from the amino acid property profiles (prop.prof.) and the amino acid distribution profiles (aa distr.) with the window size 3 as described in the text, and the Transmembrane Domain Boundaries determined by Baldwin are listed. The numbers are the positions in the graphs (the odd transmembrane sequence stretches are reversed numbered). (B) The middle points of the Transmembrane Domains, determined by three procedures, notably by Baldwin, by averaging the Transmembrane Domain Boundaries of (A), and by taking the middle point of the Transmembrane Domain Boundaries elucidated from the amino acid property profiles with window size 17. Also the deviations with the Baldwin middle points are listed (a negative value is a shift to the extracellular side of the membrane and a positive shift is a shift to the intracellular side of the membrane).

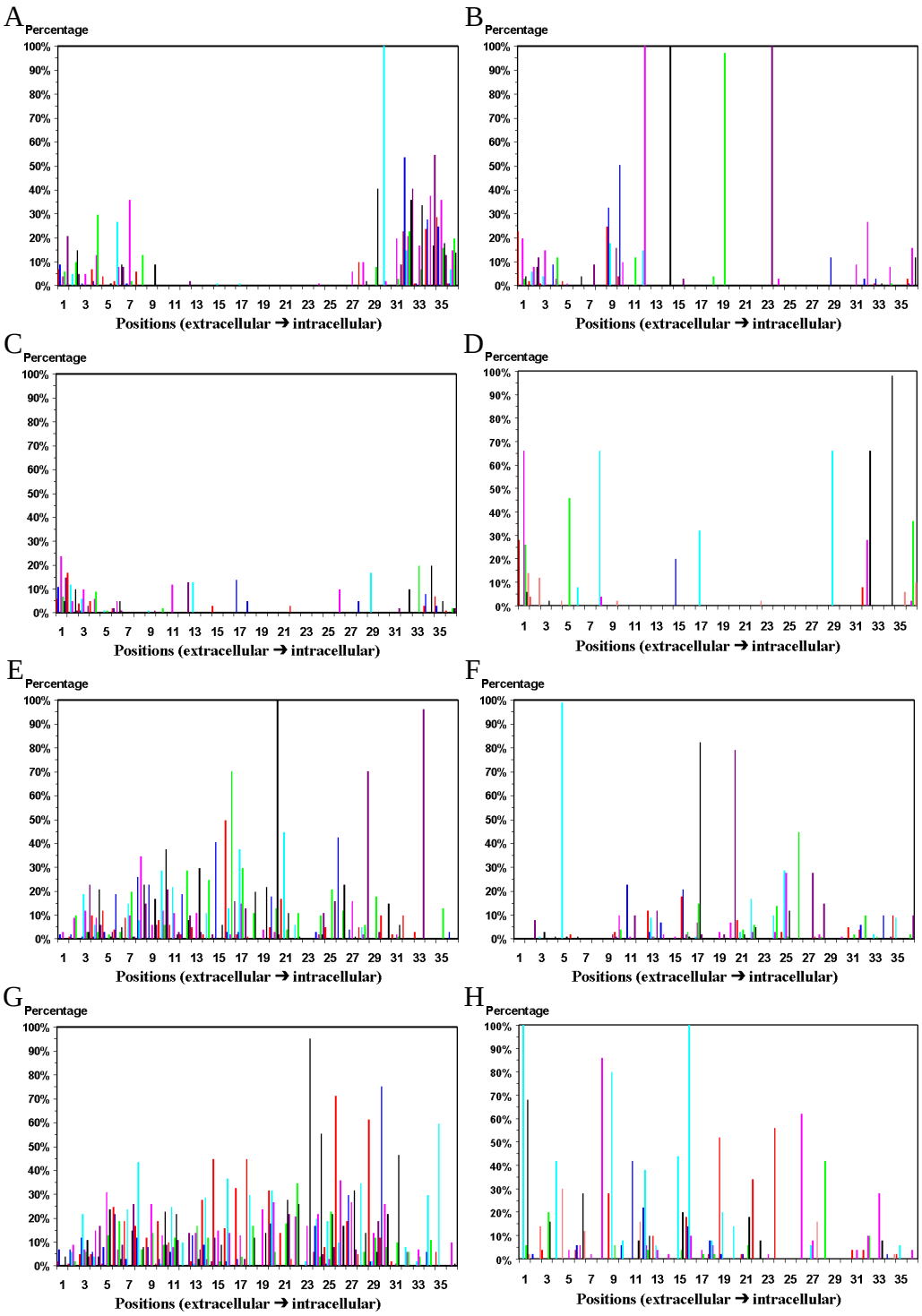


Figure 3.4: Several amino acid distribution profiles of the Transmembrane Sequence Stretches of the multiple alignment of the peptide binding GPCRs and of the multiple alignment of the opsine receptors. In each diagram the percentage of occurrence of an amino acid versus the position within the transmembrane domain is depicted with a balk. Thereby represent the different colours the seven different TMDs (1 = red, 2 = dark blue, 3 = light blue, 4 = pink, 5 = green, 6 = black, 7 = purple). The odd numbered TMDs are aligned antiparalel to the even numbered TMDs. The distribution profiles are: A. Arginine peptide GPCRs, B. Proline peptide GPCRs, C. Glutamate peptide GPCRs, D. Glutamate opsine receptors, E. Phenylalanine peptide GPCRs, F. Cysteine peptide GPCRs, G. Valine peptide GPCRs, and H. Glycine opsine receptors.

Charged residues are not necessarily excluded from the TMDs. Arginine appears with a low occurrence in the TMDs of the peptide binding GPCRs (Figure 3.4A). Lysine and aspartate occur both with a very high percentage at a number of distinct places in a few TMDs (TMD 7 of the opsine receptors and in TMD 2 of both families respectively).
The polar residues asparagine and glutamine have similar patterns of occurrence when compared to the negatively charged residues, although they seem to appear preferentially closer to the central region of TMSSs.
Proline residues can be found in transmembrane helices, accompanied with a high degree of conservation (see von Heijne [97, 109] and Williams [96]). Four strictly conserved prolines are found in the peptide binding GPCRs, in TMSS 4, 5, 6 and 7 respectively, and seven occur in the opsine receptors. Less conserved proline residues are found mostly at the extracellular side of the TMSSs (Figure 3.4B), which was also observed by Landolt-Marticorena [98].
The aromatic residues tryptophan, histidine, tyrosine and phenylalanine do not share a common occurrence pattern. The profile of histidine is similar to that of charged residues, i.e., histidine residues occur mostly at the termini of the TMDs, but they seem to have no preference for a particular side. Also tryptophan residues are found at the terminal regions. However they are found a few positions more to the central region of the TMDs, as is expected for the typical membrane boundary residue. In both receptor families, tryptophan occurs at two positions with a percentage of occurrence > 95%, notably in TMD 4 and TMD 6. The opsine receptors have a third very conserved (80%) tryptophan in TMD 3.
In contrast to tryptophan and histidine, tyrosine and especially phenylalanine frequently occur in the TMDs (Figure 3.4E). In the peptide binding GPCRs four positions were identified with a very high percentage of occurrence for tyrosine residues, namely in TMSS 5 and 7 at the intracellular border of the TMD, in TMSS 3 at the intracellular side (from the well know DRY-motif [79]) and in TMSS 1 at the extracellular border of the TMD, respectively.
Landolt-Marticorena [98] observed that phenylalanine was located within the TMD and in the intracellular flanking region. In contrast, the occurrence profiles discussed here clearly showed that they are found preferentially in the TMDs and in the extracellular flanking regions of both families (Figure 3.4E).
In the peptide binding GPCRs, cysteine residues seem to have a preference for the TMD (Figure 3.4F). Two positions have a percentage of occurrence > 75% (TMD 6 and TMD 7). The most conserved cysteine residue is found in the extracellular region of TMSS 3 with a percentage of occurrence of 99%. The occurrence of cysteine in the opsine receptors is slightly different in that more cysteine residues seem to occur. Also the strongly conserved cysteine residue in the extracellular flanking region of TMSS 3 is found, but those in TMD 6 and 7 are significantly less conserved, 56% instead of 82% and 20% instead of 79% respectively.
In contrast to cysteine, methionine seems to be randomly distributed over the TMSSs of the peptide binding GPCRs. They have a low occurrence in all the TMSSs, only three positions have a percentage of occurrence higher than 40%. Methione has a higher occurrence in the opsine receptors, but again, the distribution seems to be random over the TMSSs.
The distribution of the hydrophobic residues valine, leucine and isoleucine are not significantly different. They have a preference for the TMDs, but they are certainly not excluded from the flanking regions. According to Landolt-Marticorena [98], these residues have also a non-random distribution within the TMD. Along a path from the extracellular to the intracellular side an isoleucine, a valine and a leucine-enriched region is analysed. Such a distribution could not be observed in the occurrence profiles of these amino acids in both GPCR families investigated

here. The distribution of the valine residues in the peptide binding GPCRs is shown in Figure 3.4G.
The occurrence profiles of alanine, serine and threonine residues reveal a random distribution. In contrast, the occurrence of glycine is not entirely random, as mentioned by Landolt-Marticorena [98], these residues have a low percentage of occurrence in the intracellular regions of the TMSSs. This is especially the case for the opsine GPCRs (Figure 3.4H).

3.3.4 Determination of the Transmembrane Domain Boundaries

By combining the amino acid distribution profiles and with the amino acid property profiles, a more precise determination of the TMDBs can be made, since the distribution of certain amino acids is clearly depending on the membrane topology. The TMDBs, listed in Table 3.3A, are adapted by including the amino acid distribution profiles of the peptide binding GPCRs. The following rules were derived for this determination:

- Arginine and lysine residues are preferable located outside the TMD. When the TMD is found too short, arginine and lysine can be found on the first or the last three positions of the TMD, since the long side chain can extent over one helical turn so that the hydrophilic head group reaches the hydrophilic environment of the lipid headgroups.
- The same rule accounts for the glutamate residues.
- Since aspartate has a considerably shorter sidechain, only the first start and the terminal two positions of a TMD are accessible for aspartate.
- Tryptophane residues are situated close to the borders of the TMDs, displaying a clear preference for the membrane boundary region as was found for other membrane proteins.
- Non-highly-conserved proline residues can be found on the first three positions at the amino-terminal side and at the last position of the carboxy-terminal side of a transmembrane helix, due to the helix-breaking capability of this residue [79].

3.3.4.1 TMD 1

In the following, characteristic sequential details are introduced for each putative transmembrane helix, extracted from the multiple sequence alignment and the fingerprint of Baldwin [60, 6, 66]. In TMD 1 the positions 9 and 10 are occupied in several instances by proline residues (Figure 3.5A). The human Bradykinin-2 receptor shows prolines on both positions. Since these prolines are not widely conserved, are they estimated as transmembrane helix terminators. These positions are on the N-terminal side, therefore the first position of the TMD is position 8. However on position 8 the occurrence of arginine is 6%, that of tryptophan is 16.8%, rendering that position still being extracellular. From that finding, the first TMD position is adapted by the residue in position 9. For the intracellular end of TMD 1, a similar observation can be made. Position 28 is occupied by 16.8% tryptophan residues, 2% lysine residues, and 10% arginine residues. From this fact it can be concluded that the TMD ends at position 27. The final length of TMD 1 appears to be 19 amino acid residues, ranging from position 9 to 27.

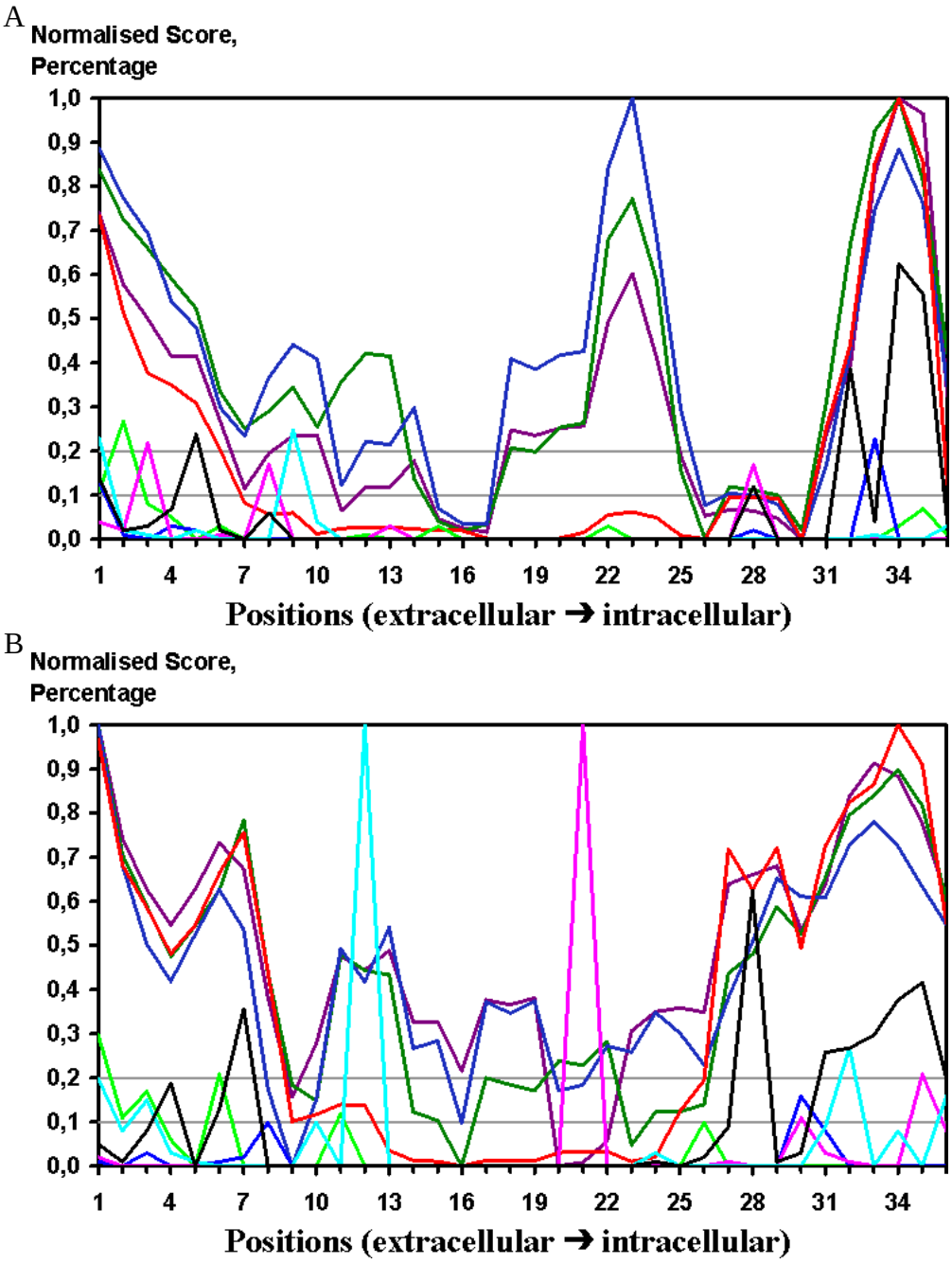


Figure 3.5: With the combination of the property profiles of the polarity scale of Zimmerman and the hydrophobicity scales of Kyte-Doolittle (dark blue), Sweet (dark green), and Hopp-Woods (purple) and the distribution profiles of the tryptophan (magenta), the arginine/lysine (black), the aspartate/glutamate (light green), the histidine (blue) and the proline (light blue) residues, the boundaries of the transmembranes are determined. In the two diagrams the normalised scores of the different property profiles and the percentage of occurrence (divided by 100) of the different amino acids are depicted for the transmembrane sequence stretches 1 (A) and 4 (B).

3.3.4.2 TMD 2

In TMD 2, the positions 9 and 10 exhibit proline residues with an occurrence of 32.7% and 50.5% respectively. Consequently, position 10 is the C-terminal end of TMD 2; corroborated by the occurrence of 16.8% tryptophan residues on position 9. The terminal residue on the opposite end of TMD 2 is position 27, since 12.9% aspartate and 5% glutamate residues are found in position 28, thus indicating a typical membrane boundary property. TMD 2 is obviously composed of a stretch of 18 residues, i.e. 10-27.

3.3.4.3 TMD 3

The identification of TMD 3 with its precise boundaries appears to be problematic. On positions 9 and 12 proline residues appear with 17.8% and 14.9% occurrence, respectively. Since this describes the N-terminal boundary, position 10 can be considered as the first transmembrane residue. However, on position 12 and 13 charged residues occur, i.e., aspartate (16.8%) and lysine (11.9%) at position 12, glutamate (12.9%) and lysine (10.9%) at position 13, respectively. From this observation, position 14 emerges as likely terminus of TMD 3, thus leading to a sequence stretch that would be too short to span the membrane. Therefore, the exception rules mentioned above are employed and position 11 is assigned as the N-terminal residue of TMD 3. According to the amino acid property profiles, position 27 appears as the last residue of TMD 3. Sequence analysis reveals position 28 as not being occupied by charged residues, tryptophan, or proline, which would indicate TMD termination. Position 29 is the first position where such residues occur, namely aspartate (81.2%), glutamate (16.8%), respectively. According to the amino acid distribution profiles, position 28 is assigned as C-terminal TMD 3 residue.

3.3.4.4 TMD 4

Similarly to the determination of TMD 3, the analysis of TMD 4 appears difficult. Even though proline is found at position 10 with an occurrence of 10%, position 11 is occupied by glutamate in 11.9% of all cases. The first observation (Figure 3.5B) suggests the residue at position 10 as the terminal TMD residue, while the second finding indicates the residue at position 12 as the boundary position. The opposite extracellular interface is defined by position 25 since glutamate is found in 9.9% and lysine in 2% of all cases at position 26. The resulting helix ranging from residue 12 to 25 would be too short to span the bilayer membrane, thus charged residues are obviously part of the transmembrane helix in its boundary regions. In position 27, arginine occurs with 6%, lysine with 2.9%, while position 28 is occupied by arginine with an occurrence of 10% and by lysine in 52.4% of all cases, respectively. Due to the high percentage of occurrence of charged residues in position 28, the preceeding position 27 is assigned as the last residue of TMD 4. This is further supported by the occurrence of tryptophan, a typical boundary residue, in this very position, although it is with a low percentage. Thorough analysis of the three hydrophobicity profiles and the polarity profile reveals position 26 as the boundary residue. This because the polarity scale of Zimmerman and the hydrophobicity scale of Kyte-Doolittle are at position 26 below the threshold value of 0.2 and at position 27 they have a significant higher value. This rises in the score from position 26 to 27 is also observed for the two other hydrophobicity scales.
Ranging from position 10 to 26, the TMD 4 helix has an extension of 17 residues.

3.3.4.5 TMD 5 and 6

The determination of the TMDBs of TMD 5 and 6 are obtained straightforward. The distribution of indicative residues such as arginine (position 11), lysine (position 29), tryptophan (position 10) and proline (position 11) results in the assignment of TMD 5 covering the sequence range from position 12 to 28. Similar findings allow the identification of TMD 6, encompassing residues between the positions 10 and 27.

3.3.4.6 TMD 7

The boundary determination of TMD 7 requires a more detailed analysis. On positions 9 and 12 aspartate and glutamate residues are found while position 13 displays tryptophan with enhanced probability. At the intracellular side, the first charged residues are found in position 31, tryptophan on position 28, respectively. Taking tryptophan as main indication for the boundary region, positions 14 and 27 define the extension of TMD 7, which results in a sequence stretch with insufficient length to span the membrane in an α-helical conformation. Taking into account that Cα-Cβ bond vectors of residues adopting an α-helical conformation are oriented towards the N-terminus, it can be assumed that a tryptophan residue at position 13 is part of TMD 7 while still exposing its sidechain into the extracellular aqueous compartment. In contrast, a tryptophan at position 28 has to be considered as boundary residue. Based on this assumption, TMD 7 comprises residues from position 12 to 28.

In summary, while comparing the TMDBs determined by the described sequence analysis tools with those published by Baldwin [66], the maximal derivation accounts for only one single sequence position, as is shown in the last column of Table 3.3.

3.3.5 Helical Wheel Analyses

In the previous section the results of thorough bioinformatical analysis of multiple GPCR sequence alignments were described, resulting in a type of a generic 1D typology of the receptor proteins in that the linear protein sequence was partitioned into seven TMDs and connecting loop regions. In the following, these results are transformed into a 2D representation by conducting extensive helical wheel analyses. For the seven TMDs, we calculated the sequence property moments following a consensus approach by applying 8 distinct property scales from Table 3.1, notably: the hydrophobicity scales of Abraham, Eisenberg, Kyte-Doolittle, Roseman, Sweet and Wolfenden, the hydropathy scale of Hopp-Woods, and the polarity scale of Zimmerman. The sequence property moments were calculated for (i) the multiple alignments of the TMDs of the separate peptide binding GPCR families, (ii) the multiple alignment of all the 101 sequences of peptide binding GPCRs, (iii) the multiple alignment of 50 opsine receptors, and (iv) the multiple alignment of 623 rhodopsin-like GPCRs. The property moment vectors were drawn in a 2D graph according to the following conventions:

- the first residue of every TMD is aligned at 0° (this is at the +Y-axis) and the central helix axis is at the origin
- helices 1, 3, 5 and 7 proceed clockwise through the membrane seen from the extracellular side
- helices 2, 4 and 6 proceed counter clockwise through the membrane seen from the extracellular side
- the second residue of an odd numbered TMD is at +100°
- the second residue of an even numbered TMD is at -100°
- the third residue is at +200° and 200°, respectively, etc.
- each residue covers an angular region of 100°, thus accounting for an ideal α-helical conformation

In an ideal arrangement of the seven α-helices the following points would be met (see Figure 3.6):

- every TMD is a perfect α-helix with a periodicity angle of 100°
- the seven TMDs have a perfect circular arrangement
- the axis of every α-helix is perpendicular to the membrane plane and perfect parallel or anti-parallel to the other α-helices
- every TMD is an ideal amphiphilic α-helix with its vectors for polarity and hydropathy pointing to the exact centre of the protein bundle and the vector for hydrophobicity orientated anti-parallel to the inside pointing vectors

Property scale	Rec. family	HELIX 1		HELIX 2		HELIX 3		HELIX 4		HELIX 5		HELIX 6		HELIX 7	
		angle	length	angle	length	angle	length	angle	length	angle	length	angle	length	angle	length
Zimmerman	peptide rec.	21.1	23.0	0.1	2727.9	56.0	317.6	36.5	132.9	-31.8	66.4	141.0	436.5	-5.2	375.2
	Opsine rec.	15.9	15.1	18.9	1668.3	170.3	354.1	60.9	94.5	-35.9	461.5	-123.5	8.2	53.4	1921.1
	623 seq.	26.4	19.3	3.9	1872.0	27.3	217.5	29.6	71.3	-22.2	47.5	155.8	69.6	-2.9	321.3
HoppWoods	peptide rec.	3.9	2.3	-13.4	15.3	21.1	16.0	103.0	2.5	-50.0	9.5	-26.9	7.0	-43.4	18.1
	Opsine rec.	23.0	8.4	27.0	37.7	42.5	24.3	54.1	0.8	-47.5	8.2	-15.2	14.0	14.8	32.6
	623 seq.	15.5	3.5	-5.5	17.8	22.2	12.2	91.7	1.1	-47.0	7.1	-14.4	6.1	-29.4	19.5
KyteDoolittle	peptide rec.	-149.3	118.8	-122.5	43.3	-148.5	17.2	130.0	19.4	-161.2	92.5	5.7	51.1	155.6	284.2
	Opsine rec.	-168.8	62.2	-110.0	53.6	-94.6	10.6	-137.2	72.5	178.4	33.7	43.7	80.0	158.1	126.7
	623 seq.	-153.0	66.9	-131.4	20.4	-155.7	9.4	148.0	31.2	-166.2	34.6	1.4	24.9	160.6	229.1
Abraham	peptide rec.	-147.8	13.3	-97.4	3.4	-169.7	9.3	140.6	4.2	151.2	5.3	47.7	1.3	135.2	43.0
	Opsine rec.	-162.3	28.9	-125.0	16.0	-143.2	21.9	-174.9	2.4	122.2	7.3	114.3	2.0	166.0	30.3
	623 seq.	-154.2	14.5	-140.9	1.3	-160.3	5.7	145.1	2.6	138.9	4.2	-41.1	0.0	143.9	33.6
Sweet	peptide rec.	-175.0	2.4	151.6	2.7	-171.6	7.7	160.9	0.9	133.5	7.2	150.0	0.4	144.2	13.7
	Opsine rec.	-161.7	6.6	-146.5	13.7	-133.5	14.9	-171.6	6.6	125.8	7.6	172.0	2.3	174.7	25.8
	623 seq.	-162.6	3.4	169.8	4.1	-161.7	4.1	163.1	2.3	127.7	6.4	-174.9	1.7	159.2	12.4
Eisenberg	peptide rec.	-136.5	5.0	-133.6	2.4	-164.9	2.7	132.4	0.5	-174.2	3.0	26.8	1.2	146.5	13.5
	Opsine rec.	-169.2	5.4	-124.9	4.3	-115.8	0.9	-148.0	1.8	159.2	3.2	62.9	2.5	168.9	8.3
	623 seq.	-146.6	3.7	-145.2	1.3	-155.9	1.3	148.5	0.7	178.4	1.6	27.7	0.7	152.9	11.9
Roseman	peptide rec.	-164.5	20.8	-165.7	10.6	-162.6	6.3	155.4	1.9	161.5	9.0	0.0	4.9	150.5	38.9
	Opsine rec.	-173.6	21.2	-141.6	16.8	-71.7	2.8	-174.7	7.4	142.7	6.2	46.8	3.1	177.2	32.0
	623 seq.	-162.4	15.7	-174.2	13.6	-136.5	3.0	152.3	3.7	151.4	2.7	36.8	0.4	154.0	34.8
Wolfenden	peptide rec.	-123.5	323.4	-113.9	229.5	-159.3	35.9	155.0	220.7	-138.0	142.4	-4.6	246.7	154.2	510.8
	Opsine rec.	-167.1	199.0	-91.3	95.2	20.8	156.1	-151.6	125.1	-177.0	47.8	27.6	227.7	136.7	93.4
	623 seq.	-135.4	201.0	-113.4	76.4	-156.3	15.6	155.8	156.6	-141.4	48.3	-1.5	166.4	155.3	360.4
Baldwin	Variability	-175.7	3326.6	-150.0	1543.0	-137.0	165.9	164.6	2100.9	-142.1	4760.9	7.5	1811.4	166.1	189.4

Table 3.4: The angles and the lengths of the vectors representing the sequence property moments of the multiple alignments of the 101 sequences of the peptide binding GPCRs, the sequence property moments of the multiple alignments of the 50 sequences of Opsine receptors and the sequence property moments of the multiple alignments of the 623 sequences of Rhodopsin-like GPCRs. The angles and the lengths of the vectors representing the variability moments of Baldwin are also listed.

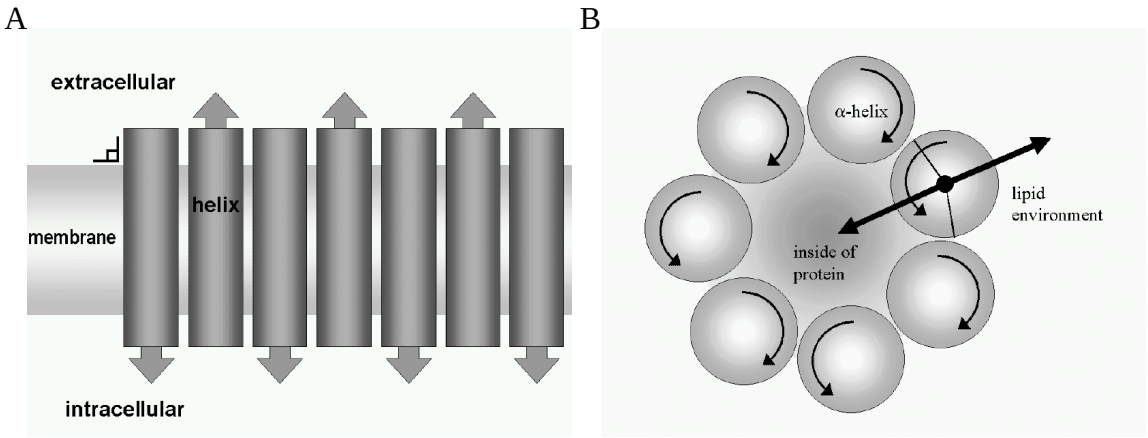


Figure 3.6: An ideal arrangement of the seven α-helices. All α-helices are perpendicular to the membrane plane (A: horizontally and perpendicular to the paper; B: the plane of the paper) and parallel or anti-parallel to each other. The inward pointing property moment is pointing to the centre of the protein bundle and the outward pointing property moment is anti-parallel to this property moment and is in the middle of hydrophobic side of the α-helix.

The angles and lengths of all vectors of the three comprehensive multiple alignments are listed in Table 3.4, together with the angles and lengths of the vectors determined with the variability determined by Baldwin [66, Figure 2 therein]. Table 3.5 gives the averaged angular orientations of the vectors from Table 3.4 clustered in two groups; notably, the group of the vectors expected to point inside the protein bundle (the vectors of the polarity scale of Zimmerman and of the hydropathy scale of Hopp-Woods) and the group of vectors expected to point to the lipid environment (the vectors of the six hydrophobicity scales). The averages were calculated with normalised vector lengths in order to compensate for the high vector lengths of the polarity scale of Zimmerman and the hydrophobicity scale of Wolfenden, respectively.

3.3.5.1 TMD 1 and TMD 2

For TMD 1 the vectors of the hydrophobicity scales have a very similar angular orientation with an average at -154.4°, thus being in close proximity to the variability vector of Baldwin (-175.7°).
The angles of the hydrophobicity vectors of TMD 2 vary between -91.3° and +169.8°. In general, the vector lengths of the hydrophobicity scales are smaller compared to those of TMD 1, however, the vector lengths of the Zimmerman and the Hopp-Wood scales are significantly larger. The deviation from an antiparallel alignment of the inside vectors and the outside vectors is decreased when compared to TMD 1, the total average of all vector graphs results in an angular orientation of 141.6°. Again, the variability vector of Baldwin (-150.0°) is close to the hydrophobicity vectors.

3.3.5.2 TMD 3

Figure 3.7 illustrates that the hydrophobicity vectors of the different peptide binding receptor families for TMD 3 based on, e.g., the Wolfenden and the Kyte-Doolittle scales hardly show a preferred orientation. However, the vectors computed from the total group of peptide binding

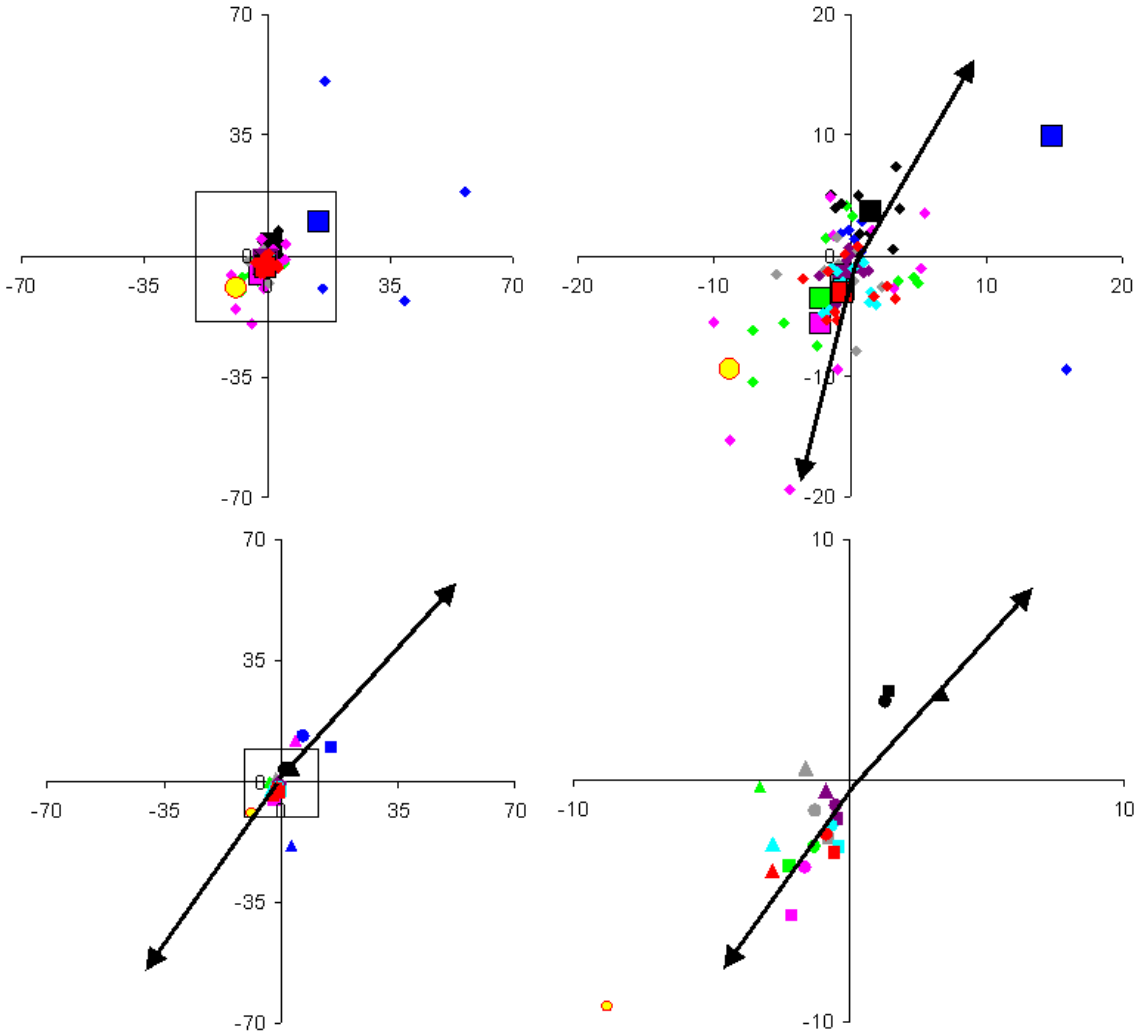


Figure 3.7: The vectors representing the sequence property moments of the transmembrane domain 3. In the first upper row the vectors of the different families of peptide binding GPCRs (small symbols) and of the total multiple alignment of the peptide binding GPCRs (large squares) are depicted. In the second row the vectors of the three major multiple alignments are shown, notably, the peptide binding GPCRs (square), the opsine receptors (triangle) and the 623 rhodopsin-like GPCRs (dot). The right part is always a blow-up of the left part. Blue = polarity of Zimmerman, green = hydrophobicity of Kyte-Doolittle, black = hydropathy of Hopp-Woods, magenta = hydrophobicity of Wolfenden, purple = hydrophobicity of Eisenberg, light blue = hydrophobicity of Sweet, red = hydrophobicity of Abraham, and yellow-red = variability of Baldwin.

	Vector	Helix 1	Helix 2	Helix 3	Helix 4	Helix 5	Helix 6	Helix 7
Av. Angle	Zimm.+ HW all	17.7	5.1	43.2	62.3	-39.1	-52.4	-3.0
	Zimm.+ HW pept.	12.5	-6.6	38.6	69.7	-40.9	57.0	-24.3
	Zimm.+ HW ops.	19.5	22.9	106.4	57.5	-41.7	-69.3	34.1
	Zimm.+ HW 623 seq.	21.0	-0.8	24.8	60.7	-34.6	70.7	-16.1
	Hydro. All	-156.4	-136.5	-145.4	165.3	164.7	34.6	155.2
	Hydro. Pept. GPCRs	-149.4	-138.0	-162.8	145.7	175.3	24.8	147.7
	Hydro. Opsine	-167.1	-123.4	-101.2	-159.7	151.8	70.6	163.7
	Hydro. 623 seq.	-152.4	-148.9	-154.5	152.1	167.7	5.3	154.3
Δ Angle	All	174.0	141.6	171.4	103.0	156.3	87.0	158.2
	Peptide receptors	161.9	131.4	158.7	76.0	143.9	32.2	172.0
	Opsine receptors	173.4	146.3	152.4	142.8	166.5	140.0	129.6
	623 sequences	173.3	148.1	179.2	91.5	157.8	65.5	170.5

Table 3.5: The average angles of the vectors are expected to point inside the protein bundle (these are the vectors of the polarity scale of Zimmerman and of the hydropathy scale of Hopp-Woods) and the average angles of the vectors are expected to point to the lipid environment (the vectors of the hydrophobicity scales). These angles are listed for the three major multiple alignments; notably, the peptide binding GPCRs, the opsine receptors and the 623 rhodopsin-like GPCRs. For comparison, also the smallest angles between the averages of the inside and the outside pointing vectors are listed.

GPCRs have similar angles for the different hydrophobicity scales, but the vector lengths are small when compared to those of other TMDs. The property moments calculated from the 623 aligned sequences are in close proximity of those originating from the peptide binding GPCR alignment (Figure 3.7). The property moments derived from the opsine receptor alignment show great discrepancies. The vectors of the Wolfenden, Kyte-Doolittle, Roseman and Zimmerman scales are spread over the entire graph encompassing a variety of different angles (Table 3.4). The average angle of the hydrophobicity vectors for the opsine receptors is 101.2°, whereas it is 162.8° and 154.5° for the peptide binding GPCRs and the 623 rhodopsin-like sequences, respectively. The variability of Baldwin results in the smallest vector for this TMD, but its direction is similar to that of the hydrophobicity moments of the total peptide binding GPCR and rhodopsin-like GPCR alignments (-137.0° versus 162.8° and 154.5°, respectively). The Hopp-Woods scale derived vectors are in close agreement among the different GPCR sequence alignments.
The differences in the vectors of the Hopp-Woods scale are in contrast rather small, 42.5° versus 21.1° and 22.2°. They cannot be pinpointed to one residue which is responsible for these differences. However, the most significant differences between the peptide binding GPCRs and the opsine receptors in this TMD, is that the opsine receptors have a moderately conserved glutamate residue on position 17 and that on position 12 of the peptide binding GPCRs there are different polar residues found, whereas on the same position in the opsine receptors no polar residues are present.

3.3.5.3 TMD 4

Within TMD 4, the hydrophobicity moments are pointing in very similar directions, the variation is found between -137.2° to 130.0°. The angle of the variability moment of Baldwin is again in the area of the hydrophobicity based vectors (164.6° versus -159.7° - opsins; 145.7°

- peptide binding GPCRs; 152.1° - rhodopsin-like GPCRs). The differences between the three groups for the inward oriented moments is not that significant.

Interestingly, the most important observation for this TMD is that the angle between the averages of the inside- and outside-pointing moments is deviating dramatically from an antiparallel orientation (180.0°), notably 76.0° for the peptide-binding GPCRs, and 91.5 for the rhodopsin-like GPCRs, respectively.

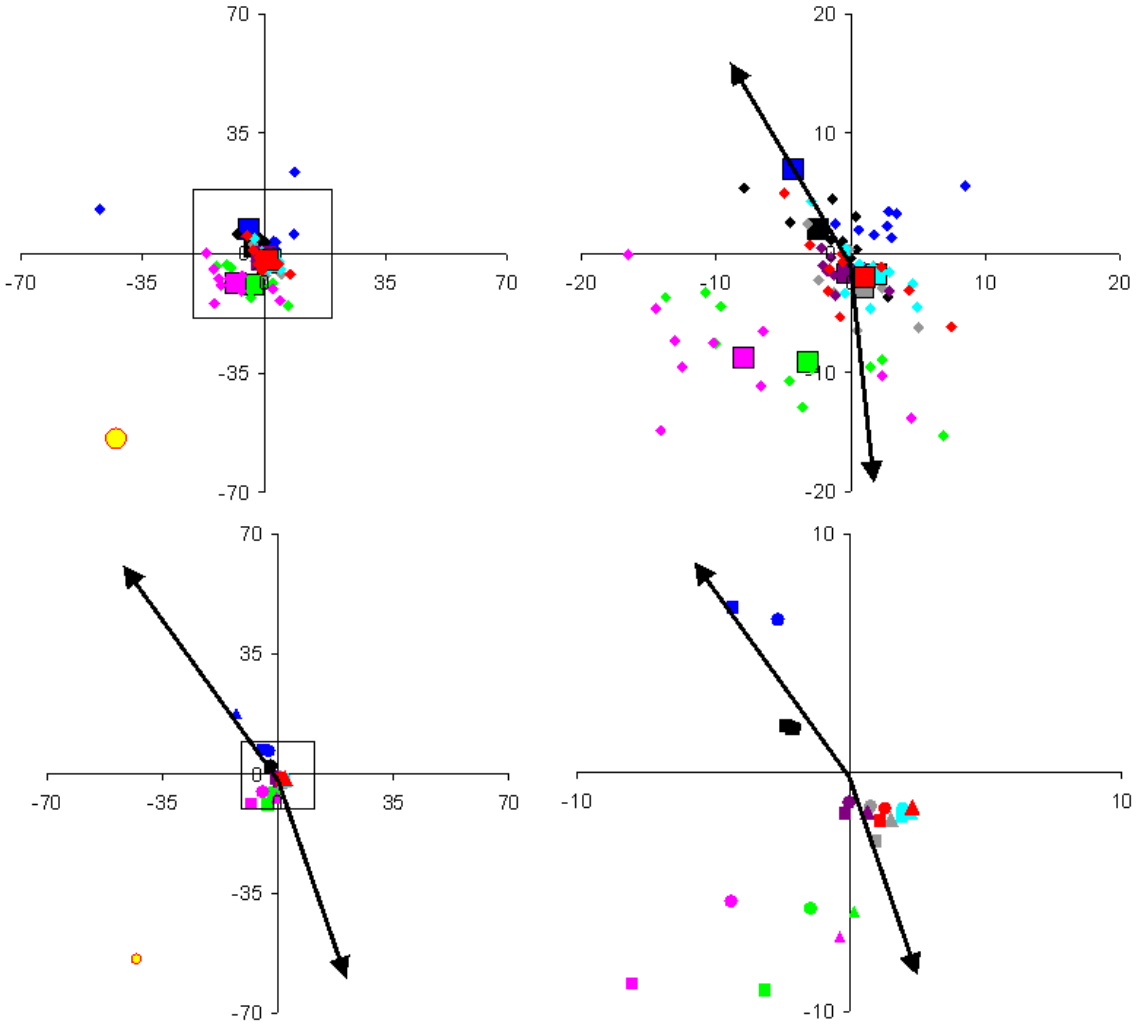


Figure 3.8: The vectors representing the sequence property moments of the transmembrane domain 5. In the first upper row the vectors of the different families of peptide binding GPCRs (small symbols) and of the total multiple alignment of the peptide binding GPCRs (large squares) are depicted. In the second row the vectors of the three extensive multiple alignments are shown, notably, the peptide binding GPCRs (square), the opsine receptors (triangle) and the 623 rhodopsin-like GPCRs (dot). The right part is always a blow up of the left part. Colours see Figure 3.7.

3.3.5.4 TMD 5

Figure 3.8 depicts the vectors computed for the eight property scales based on 11 different multiple alignments of the peptide binding GPCRs of TMD 5. The vectors of the different receptor families (small symbols) display severe variation in orientation, however, not significantly different compared to the other TMDs. In contrast, the vectors of the three comprehensive multiple alignments, comprising the 101 peptide binding GPCRs, the 50 opsine receptors, and the 623 rhodopsin-like GPCRs show similar lengths and orientations. It is further noteworthy that vector pairs, expected to be aligned in an antiparallel orientation, such as hydrophobicity/polarity vectors, or Baldwin variability/hydropathy vectors encompass an average angle of 155.0°.

3.3.5.5 TMD 6

As found for TMD 3, some hydrophobicity moments of the different peptide binding receptor families are widely spread over the graph for TMD 6. The moments derived from the hydrophobicity scales of Abraham, Sweet, Eisenberg and Roseman display one of the smallest values when TMD 6 is compared to the other TMDs. The averages of the angles for the Kyte-Doolittle and Wolfenden moments (0.6°, 35.7° and 0.1° for the peptide binding GPCRs, the opsine receptors and the rhodopsin-like GPCRs) are similar to the angle of the variability moment of Baldwin (7.5°).
Remarkably, the moments of the Hopp-Woods scale are orientated entirely different when compared to the moments of the Zimmerman scale. The Hopp-Woods moments align well with the Kyte-Doolittle and the Wolfenden moments. The Zimmerman moments are aligned almost anti-parallel to the Kyte-Doolittle and the Wolfenden moments; 140.5°, 159.2° and 155.9° for the peptide binding GPCRs, the opsine receptors and the rhodopsin-like GPCRs, respectively.

3.3.5.6 TMD 7

The vector moments for TMD 7 are given in Figure 3.9. The two vector groups are closer to an antiparallel orientation when compared to, e.g., TMD 5 (Figure 3.9), and the angular spread of the same property scale vectors is smaller. Only the vectors of the polarity-scale of Zimmerman and the hydropathy scale of Hopp-Woods of the opsine receptors show significant discrepancies with respect to the ideal pattern. The variability moment of Baldwin is found in the same direction (166.1°) as the hydrophobicity moments, which are spread between 135.2° and 177.2°. The inside-pointing moments for the peptide binding GPCRs, the opsine receptors and the rhodopsin-like GPCRs show averaged orientations at 24.3°, 34.1° and 16.1° respectively.

In summary, the average of the sequence property moments of the inward pointing vectors and the average of the sequence property moments of the outward pointing vectors are in TMDs 1, 2, 3, 5 and 7 in a fairly anti-parallel orientation analogous to the theoretically ideal situation (Figure 3.6). Although, sometimes individual property moments do not follow the theoretical rules. The sequence property moments of TMDs 4 and 6 are deviating from the ideal situation. In the case of TMD 4 the averages of the inward and the outward pointing vectors are almost perpendicular to each other, whereas the property moments of TMD 6 have a rather large spread of angles within the two vector groups.

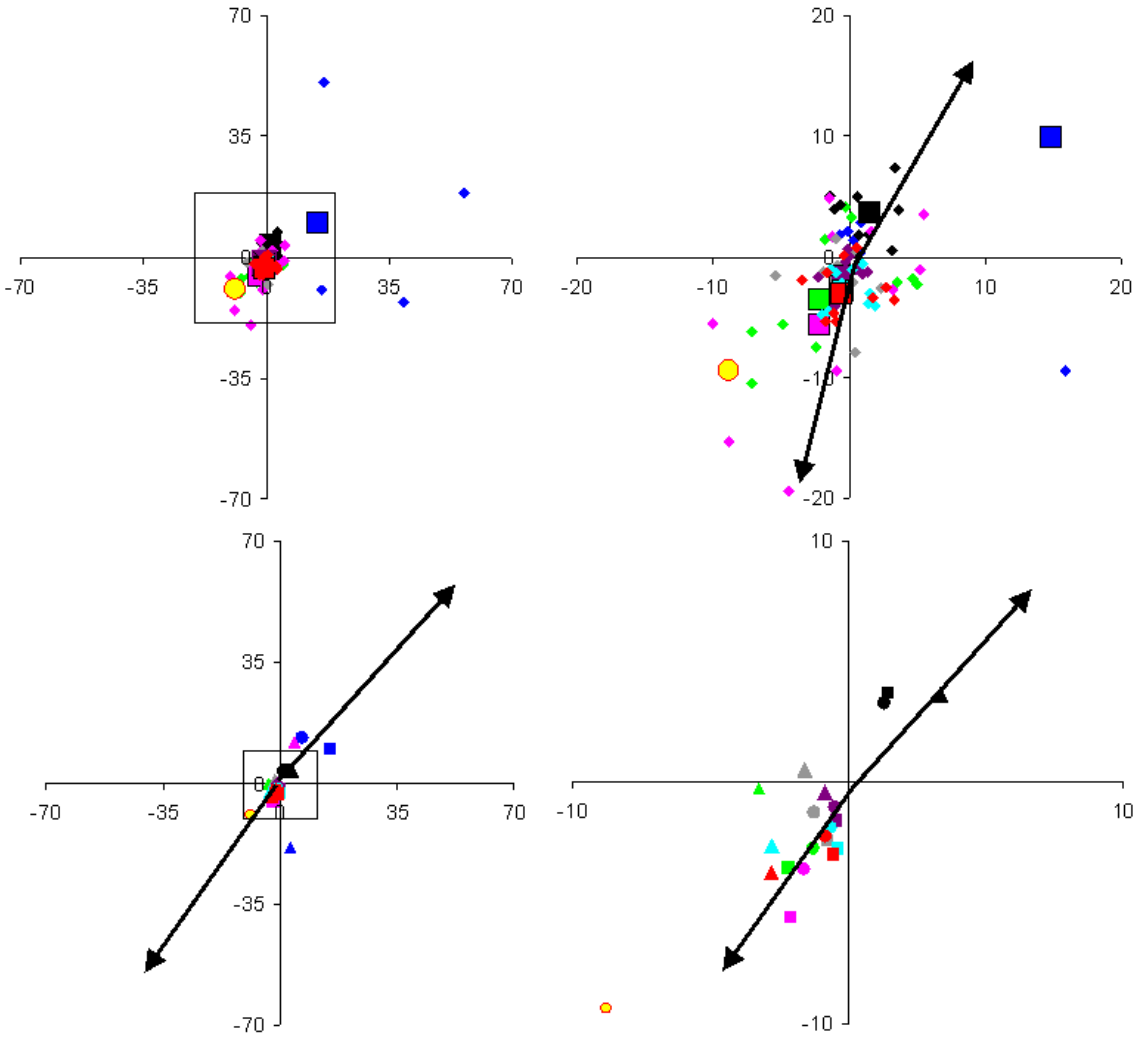


Figure 3.9: The vectors representing the sequence property moments of the transmembrane domain 7. In the first upper row the vectors of the different families of peptide binding GPCRs (small symbols) and of the total multiple alignment of the peptide binding GPCRs (large squares) are depicted. In the second row the vectors of the three extensive multiple alignments are shown, notably, the peptide binding GPCRs (square), the opsine receptors (triangle) and the 623 rhodopsin-like GPCRs (dot). The right part is always a blowup of the left part. Colours see Figure 3.7.

3.3.6 2D Projection Map of Helical Wheels

The sequence property moments are subsequently projected onto the 2D structure map of rhodopsin. Figure 3.10 illustrates this attempt for the projection of the sequence property moments computed for the peptide binding GPCRs. The variability vectors of Baldwin were positioned in the centre of the lipid bilayer at a level of 0 Å along the membrane normal of the 2D map. Since the helices are not exactly orthogonal to the membrane plane and because they are not exactly parallel to each other, the helical wheels differ significantly in the maps at +12 and -8 Å when compared to the initial 0 Å map. Due to this orientational and rotational tilt, the

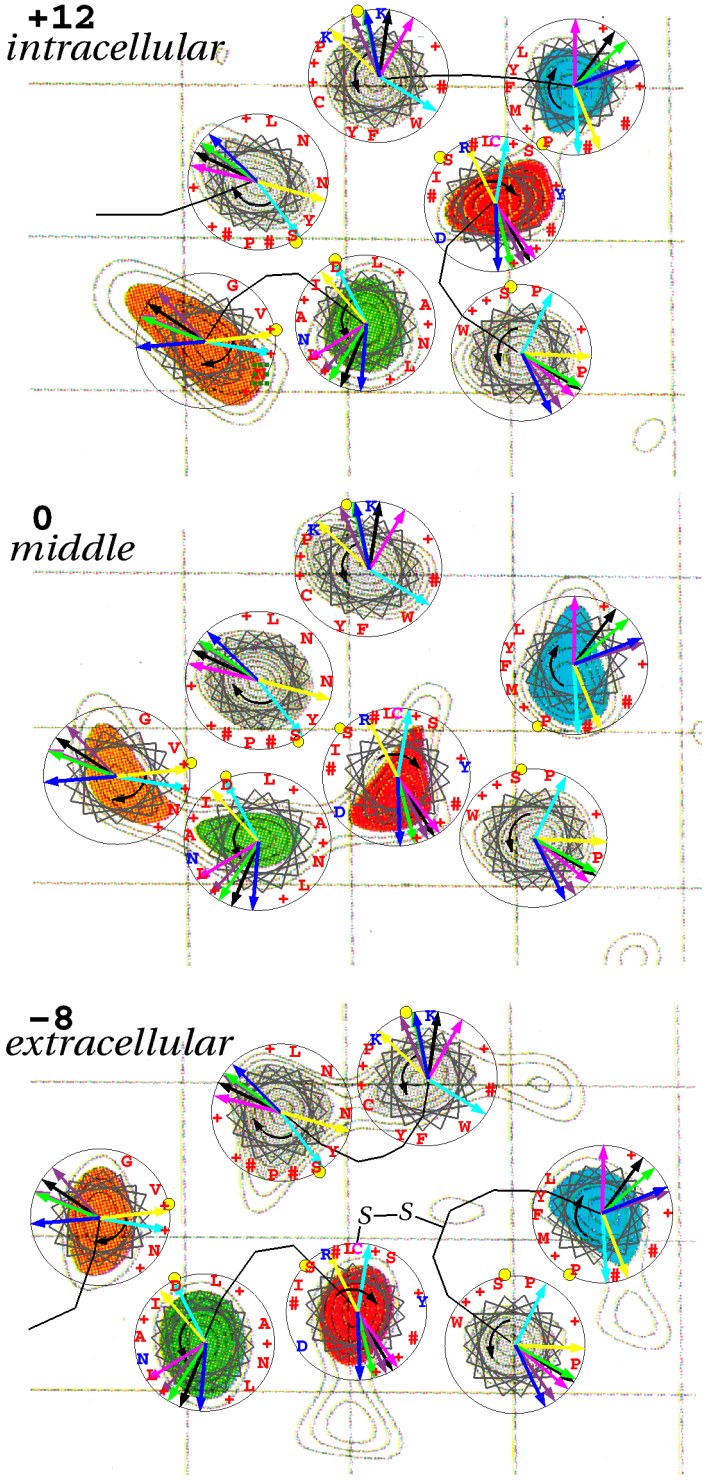


Figure 3.10: The vectors of the property moments of the peptide binding GPCRs are projected on the 2D maps of rhodopsin at the levels of +12, 0 and -8 Å, with the variability vectors of Baldwin positioned in the centre of the helical face exposed to the lipid bilayer at a level of 0 Å along the membrane normal of the 2D map. Blue = variability of Baldwin, magenta = hydrophobicity of Abraham, green = hydrophobicity of Kyte-Doolittle, black = hydrophobicity of Eisenberg, purple = hydrophobicity of Wolfenden, yellow = hydropathy of Hopp-Woods, and light blue = polarity of Zimmerman.

variability moments of Baldwin deviate from the ideal orientation when the 0 Å map of, e.g., TMD 3 (Figure 3.10) is compared with the boundary maps at +12 and -8 Å.

3.3.7 Property Moments for Proteins with Known 3D Structure

To validate our considerations, the sequence and the structural property moments were calculated for two proteins with determined high resolution 3D structures. As the closest topological analogue to GPCRs, the X-ray structure of the light-driven ion pump bacteriorhodopsin was analysed together with the four-helix-bundle protein Felix (PDB-code: 1FLX and 3FLX) [110, 111]. The later is a de-novo designed globular protein with helices adopting ideal conformations without significant kinks.
For the computation of the sequence property moments of bacteriorhodopsin, a multiple alignment of 12 sequences of the bacteriorhodopsin family was used, including, e.g., halorhodopsin [51, 52, 112] and archaerhodopsin. Combining the 3D structure of bacteriorhodopsin with the property profiles of Zimmerman (window size = 3) derived from the multiple alignments of the seven potential TMSSs of the bacteriorhodopsin-family revealed the TMDBs. According to the numbering scheme of the Brookhaven Protein Databank entry 2BRD the TMDs extend over the following seven stretches: 11-29, 44-60, 80-97, 108-125, 136-153, 173-190 and 205-222. The vectors for the sequence property moments based on the multiple alignments of the TMDs of bacteriorhodopsin were calculated for exactly the same eight scales as described above (Table 3.6).

	HELIX 1		HELIX 2		HELIX 3		HELIX 4		HELIX 5		HELIX 6		HELIX 7	
	Angle	Length	Angle	Length	Angle	Length	Angle	Length	Angle	Length	Angle	Length	Angle	Length
Zimmerman	23.6	1.8	-12.3	3.8	115.2	96.7	142.9	48.0	13.26	2.7	-18.0	46.7	140.8	94.8
HoppWoods	46.8	2.8	30.6	1.0	97.9	11.0	-160.1	3.8	149.5	2.2	-93.2	5.3	14.0	9.0
KyteDoolittle	-154.6	8.2	144.5	11.2	-47.1	8.6	57.4	9.6	-115.4	3.6	173.5	18.0	-31.6	16.5
Abraham	-136.6	4.2	-164.3	2.7	-68.3	8.6	74.6	6.2	-35.0	2.0	102.9	4.3	-24.3	8.0
Eisenberg	-149.1	2.0	151.8	2.3	-54.5	4.7	51.3	2.4	-100.2	1.2	155.9	3.6	-30.8	5.2
Roseman	-169.5	2.0	-173.1	2.9	-58.7	9.1	17.7	3.1	3.3	2.1	149.4	5.7	-19.0	8.3
Sweet	-136.7	1.4	-119.6	1.1	-83.7	4.7	60.3	3.4	-41.4	1.1	113.8	3.7	-23.8	3.8
Wolfenden	-156.7	8.4	159.1	16.2	-36.8	31.2	34.1	9.9	-174.3	8.5	178.8	30.6	-33.9	31.4
Average Zimm.+HW	35.2		9.1		106.5		171.4		81.3		-55.6		142.49	
Average Hydro.	-150.5		-180.3		-58.2		49.2		-77.2		145.7		-27.2	
Delta	174.3		170.6		164.7		122.2		158.5		158.7		169.6	
Inside 3D structure	42		-17		-139		149		0		66		160	

Table 3.6: The different sequence property moments of the bacteriorhodopsin multiple alignment, together with the average over the hydrophobicity moments, the average over the polarity plus the hydropathy moments, the delta angle between these two average moments and the angle of the middle of the inside region of the helix within the 3D-structure of bacteriorhodopsin (2BRD) are shown.

Comparable to the results obtained for the GPCRs, the property moments of TMD 1 of bacteriorhodopsin show an expected antiparallel pattern. The hydrophobicity moments have an average orientation of -150.5°, which is almost anti-parallel (174.3°) to the average orientation of the moments derived from the polarity scale of Zimmerman and the hydropathy scale of Hopp-Woods (average: 35.2°).
The hydrophobicity moments of TMD 2 cover a much larger area (144.5° to -119.6°), while the moments of Zimmerman and Hopp-Woods deviate mutually by 42.9°. However, the delta

angle between the corresponding averages accounts for an almost ideal antiparallel orientation (170.6°).

Together with TMD 1, the TMDs 3, 6, and 7 display this same anti-parallel pattern (164.7°, 158.7° and 169.6°, respectively), as is theoretical expected. The spread for the hydrophobicity vectors of TMDs 3 and 7 is moderate, (46.9° and 15.0°, respectively), but for TMD 6 it is somewhat higher (75.9°). A similar spread is also found for the inward oriented moments. The found values are 17.3°, 75.2° and 3.1° for TMDs 3, 6 and 7, respectively.

The spread of the angles of the moments of TMD 4 is moderate (56.8° and 57.0° for the hydrophobicity and the Zimmerman-Hopp-Woods moments, respectively), but the averages of these two groups are far from anti-parallel (122.2°). Therefore it can be summarised, that beside for TMD 4, the sequence property moments of the TMDs of bacteriorhodopsin show the expected anti-parallel orientation of the average of the hydrophobicity vectors versus the average of the polarity/hydropathy vectors.

The structural property moments were calculated for the 3D structure of bacteriorhodopsin taken from the Brookhaven Protein databank (entry: 2BRD) with the use of equation 3.2. As mentioned above, the unit vector for a certain residue is aligned with the Cα-Cβ bond in direction of the sidechain. In order to allow a comparison of the sequence property moment vectors with those calculated from the 3D structure, the vectors derived from the structure need to be projected into a corresponding plane. The required vector components need to be perpendicular to the main helix axis. Furthermore, the helices are rotated in such a way that the best agreement with the ideal helical wheel of the TMSSs is achieved in order to achieve a good comparison. Table 3.7 lists the lengths of these projected vectors as well as the lengths and angles of the vectors derived from the 3D structure.

Comparison of the alignment-based computed vectors with those from the 3D structure wheels reveals no uniform deviation. Differences in orientation range from, e.g., 19.5° for the Eisenberg-vector of TMD 2 up to 176.9° for the Abraham-vector of TMD 5. A comparable result is obtained for the average vector orientations. Only the deviations from antiparallel orientation of inside- and outside-pointing vectors show satisfactory correspondence, but again without any regularity.

From the 3D structure of bacteriorhodopsin the angles of the helices between the start residue of the helix and the point representing the inside of the protein were determined in order to compare these angles with the angles of the calculated property moments. These inside 3D structure angles are also listed in Tables 3.6 and 3.7. When these angles are compared with the sequence property moments, the centres of TMDs 1, 2, and 7 are rather well predicted. The average of the Zimmerman and Hopp-Woods vectors are almost parallel and the averages of the hydrophobicity vectors are almost antiparallel to the determined inside-angle. For the residual four TMDs, angles smaller than 90° can be observed.

The structural property moments predicted the centres of TMDs 2, 3, 4, 5 and 7 with acceptable accuracy, while the prediction for TMDs 1 and 6 revealed deviations of 109° and 162°, respectively. This is almost the opposite direction.

For the Felix four-helix-bundle were the sequence property moments were from a single sequence. Since the helices of Felix are almost ideal, one would expect the differences between the sequence and structural property moments to be small. In Table 3.8A, the angles and the lengths of the eight different sequence property moments are listed, while Table 3.8B gives the values for the structural property moments.

	HELIX 1			HELIX 2			HELIX 3		
	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.
Zimmerman	95.0	5.8	2.9	-59.9	7.4	1.4	-172.4	126.7	107.5
HoppWoods	-152.3	13.3	4.2	73.4	8.1	1.9	-179.7	13.4	9.2
KyteDoolittle	-92.9	19.1	5.2	92.3	10.2	4.0	46.7	14.1	10.7
Abraham	7.4	13.7	4.2	-113.6	7.6	2.6	28.4	14.9	6.8
Eisenberg	-11.3	4.9	1.2	170.9	4.0	0.4	38.9	6.5	4.6
Roseman	52.9	8.1	1.4	165.6	5.2	2.4	20.8	10.3	9.8
Sweet	22.9	8.7	2.6	-94.0	3.6	1.7	-3.8	6.1	4.4
Wolfenden	-87.2	13.6	10.7	107.2	10.2	4.4	40.1	40.5	35.2
Average Zimm.+HW	151.4			6.8			-176.1		
Average Hydro.	-18.0			174.7			28.5		
Delta	169.48			168.0			155.4		
Inside 3D-structure	42			-17			-139		

	HELIX 4			HELIX 5			HELIX 6			HELIX 7		
	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.
Zimmerman	77.6	50.6	38.2	56.9	7.2	2.4	-67.7	59.6	48.8	-144.7	101.2	82.1
HoppWoods	105.0	5.0	1.9	-33.9	13.0	1.5	-136.2	13.5	7.48	-144.7	8.5	5.4
KyteDoolittle	-30.9	14.6	2.7	174.3	14.4	3.7	123.7	21.1	16.6	16.9	19.3	10.9
Abraham	21.5	7.7	2.6	148.1	12.1	2.0	43.0	12.8	6.7	37.5	10.5	4.1
Eisenberg	-60.3	6.1	0.4	167.6	6.0	0.9	93.1	6.2	4.2	26.3	4.5	2.7
Roseman	-94.4	2.7	1.8	121.1	6.8	4.3	62.9	4.4	4.4	61.1	7.9	7.1
Sweet	1.0	2.0	1.8	176.2	6.0	2.4	56.4	5.7	3.5	50.8	6.4	2.0
Wolfenden	-77.2	12.0	5.9	-135.5	9.9	5.0	124.9	37.0	31.7	33.7	24.7	21.0
Average Zimm.+HW	91.3			11.5			-102.0			-144.7		
Average Hydro.	-40.1			168.6			84.0			37.7		
Delta	131.4			157.1			174.1			177.6		
Inside 3D-structure	149			0			66			160		

Table 3.7: The same structural property moments as in Table 3.6, but now for the 3D-structure of bacteriorhodopsin (pdb-entry 2BRD). The projected length (Proj. L.) is the length of the moment projected on the 2D helical wheel of the helix.

Due to the fact that Felix is a globular soluble protein, the hydrophobicity vectors are oriented towards the centre of the helix bundle, while the Zimmerman and Hopp-Woods vectors point to the exterior. In helix 1, the inside and outside-pointing vectors are in an almost anti-parallel orientation irrespective of the applied method for computation. However, the inside-pointing-vectors of the sequence property moments are in closer agreement with the 3D structure than those of the structural property moments. Comparable results are obtained for helix 2. The vectors of helix 2 have a remarkable small spread, which is in contrast to those of helix 3. In helix 3, the sequence and the structural property moments are oriented in entirely different directions with angular deviations of larger than 90°. In terms of accuracy, the structural property moments are more in accordance with the structural requirements referring to the orientation of opposite vector pairs. For illustration, the structural property moments for the Hopp-Woods and Kyte-Doolittle scales are depicted within the 3D structure of bacteriorhodopsin and the four-helix bundle protein Felix in Figure 3.11A and Figure 3.11B, respectively.

Helix 4 exhibits a smaller spread of their vectors, but again, a large deviation between the sequence and structural property moments in the range of 60° is found. In contrast to helix 3, the sequence property moments yield the more accurate vector pattern.

A

	HELIX 1		HELIX 2		HELIX 3		HELIX 4	
	Angle	Length	Angle	Length	Angle	Length	Angle	Length
Zimmerman	168.3	200.0	-94.1	179.3	172.6	146.7	-109.5	100.1
HoppWoods	158.2	22.5	-93.49	11.7	-164.1	8.6	-94.4	7.0
KyteDoolittle	-19.4	32.3	91.8	28.4	6.6	10.7	81.2	17.7
Abraham	-14.0	14.4	93.6	9.6	26.3	5.0	90.2	6.2
Eisenberg	-9.29	10.6	91.9	7.5	-6.9	2.6	86.6	4.5
Roseman	-9.0	19.3	91.2	13.8	11.1	5.7	90.3	8.6
Sweet	-30.9	10.1	96.5	5.9	50.4	3.8	96.1	4.7
Wolfenden	2.0	52.6	89.6	54.1	-48.9	14.8	87.0	25.7
Average Zimm.+HW	163.2		-93.8		175.8		-102.0	
Average Hydro.	-13.4		92.4		6.4		88.6	
Delta		176.7	173.8		169.4		169.5	
Inside Angle	-27		80		68		62	

B

	HELIX 1			HELIX 2			HELIX 3			HELIX 4		
	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.
Zimmerman	-158.2	220.3	169.5	-137.7	196.1	154.3	-112.6	207.0	133.1	-161.4	183.8	117.9
HoppWoods	-169.3	18.2	18.2	-138.6	8.8	8.6	-96.2	9.3	8.9	-161.4	7.1	7.0
KyteDoolittle	15.0	26.5	26.3	46.0	20.9	20.8	86.2	17.2	14.3	30.1	17.7	16.7
Abraham	19.5	12.6	11.6	47.1	6.4	6.3	99.0	6.7	6.4	20.9	5.2	5.2
Eisenberg	24.9	8.5	8.5	46.4	6.2	6.2	97.1	4.2	3.6	32.0	4.2	4.0
Roseman	24.9	16.6	15.6	47.5	11.1	10.5	98.3	11.8	9.0	36.1	10.1	8.3
Sweet	5.0	8.3	8.3	46.9	3.8	2.8	100.9	6.0	5.0	13.7	4.6	4.0
Wolfenden	36.3	64.0	49.5	46.7	56.1	50.1	105.7	57.8	24.8	20.8	53.4	30.2
Average Zimm.+HW	-163.8			-138.2			-104.4			-161.4		
Average Hydro.	20.9			46.8			97.9			25.6		
Delta	175.3			175.1			157.7			173.0		
Inside Angle	-27			80			68			62		

Table 3.8: The sequence (A) and the structural (B) property moments of the globular four-helix bundle protein Felix.

3.3.8 Property Moments for a 3D Model of a GPCR

Apart from the experimentally derived protein structures, a molecular model of a GPCR, notably the human CCK-B receptor was used as template structure for vector correlations. This model was constructed on the basis of the α-carbon trace model of Baldwin and with the TMDs determined in this study (a more detailed description of the modelling procedure will be given in Chapter 4). The angles and the lengths of the structural property moments are listed in Table 3.9. The structural property moments for the Hopp-Woods and Kyte-Doolittle scales within the 3D model of the human CCK-B receptor are depicted in Figure 3.11C.
It is obvious that the discrepancies between the sequence property moments of the peptide binding GPCRs and the structural property moments calculated for the CCK-B receptor model are in the same order as the discrepancies found for bacteriorhodopsin.

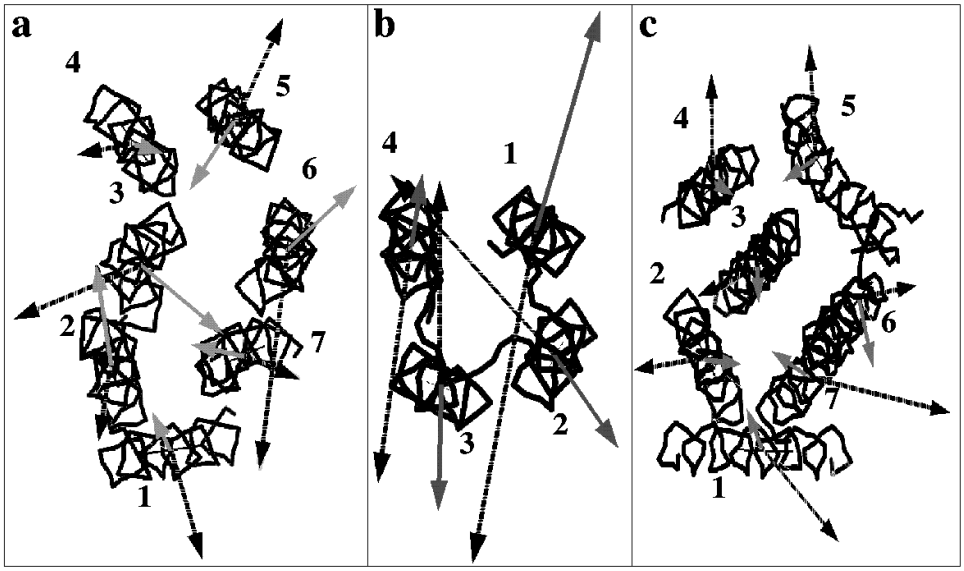


Figure 3.11: Structural property moments with the Hopp-Woods (grey) and the Kyte-Doolittle (black-stratched) scales for the 3D structure of bacteriorhodopsin (pdb-entry: 2brd) (A), for the four-helix bundle protein Felix (pdb-entry: 1flx and 3flx) (B) and for a 3D model of the human CCK-B receptor (C).

	HELIX 1			HELIX 2			HELIX 3		
	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.
Zimmerman	71.6	7.2	2.9	-27.5	52.5	43.2	-20.0	6.5	1.2
HoppWoods	22.8	9.3	2.4	12.6	8.5	2.9	32.4	6.4	1.0
KyteDoolittle	-130.0	16.3	11.9	-167.7	18.4	5.6	-118.9	13.9	3.0
Abraham	-127.6	9.3	4.3	-115.2	10.3	1.9	-134.3	5.6	0.9
Eisenberg	-129.5	5.5	2.9	-174.8	5.6	1.5	-152.1	4.5	0.9
Roseman	-110.4	5.7	3.6	107.7	5.3	1.7	0.0	3.7	1.3
Sweet	-147.5	4.5	1.4	-153.4	3.2	1.0	-23.7	1.6	0.5
Wolfenden	-111.4	19.6	18.4	165.3	10.9	9.9	167.2	7.9	5.0
Average Zimm.+HW	47.2			-7.5			6.2		
Average Hydro.	-126.1			-176.4			-103.6		
Delta	173.3			168.9			109.8		
Inside 3D-structure	70			35			111		

	HELIX 4			HELIX 5			HELIX 6			HELIX 7		
	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.	Angle	Length	Proj.L.
Zimmerman	-24.9	7.2	4.9	116.1	4.5	3.2	-114.0	5.0	1.7	80.0	60.6	43.8
HoppWoods	-177.5	10.5	1.2	-27.4	12.2	4.0	-47.7	13.8	3.9	-7.3	7.8	3.6
KyteDoolittle	143.0	17.3	9.3	-114.9	21.9	7.6	41.9	20.6	4.0	-125.7	20.8	19.5
Abraham	103.7	11.6	3.8	156.3	14.2	2.7	79.3	14.5	2.5	162.3	9.0	6.5
Eisenberg	132.9	5.7	1.8	-166.8	7.6	2.0	89.5	6.6	0.9	-133.0	4.5	3.5
Roseman	89.0	5.70	3.0	132.3	8.5	3.4	67.3	7.7	3.0	-145.1	5.7	5.4
Sweet	44.8	5.0	1.3	147.4	6.5	2.6	79.8	9.0	4.2	-154.8	3.4	3.0
Wolfenden	128.8	22.6	22.5	-92.2	15.3	11.0	-10.4	6.8	6.8	-122.2	35.5	31.8
Average Zimm.+HW	-101.2			44.4			-80.9			36.4		
Average Hydro.	107.0			-169.7			57.9			-146.4		
Delta	151.8			146.0			138.8			177.2		
Inside 3D-structure	12			20			-168			35		

Table 3.9: The structural property moments of the 3D-model of the human CCK-B-receptor.

3.3.9 TMD Determination with Perscan

Apart from the comprehensive sequence analysis study described above, an automated TMD determination with the computer program Perscan was performed with the multiple alignments of the different peptide binding GPCRs and with the comprehensive multiple alignment of the 101 sequences of peptide binding GPCRs. In order to evaluate the influence of the window size, numerous different ranges were tested: 7-19, 10-30, 10-36, 16-30, 16-36, 16-28, 18-26, 19-24, 24-28, 5-30 and 15-22. Thereby the first number corresponds to the minimum length and the second number the maximum length for a determined transmembrane sequence. A window size of 18-26 emerged as yielding the best results.
Perscan [107, 67, 108] creates a so-called power spectrum for each calculation from which the periodicity angle, the regular angle between subsequent residues, is extracted. A maximum in the power spectrum at approximately 100° accounts for the presence of a region with α-helical periodicity. Not all calculations revealed the desired peak at 100°. However, the residues per

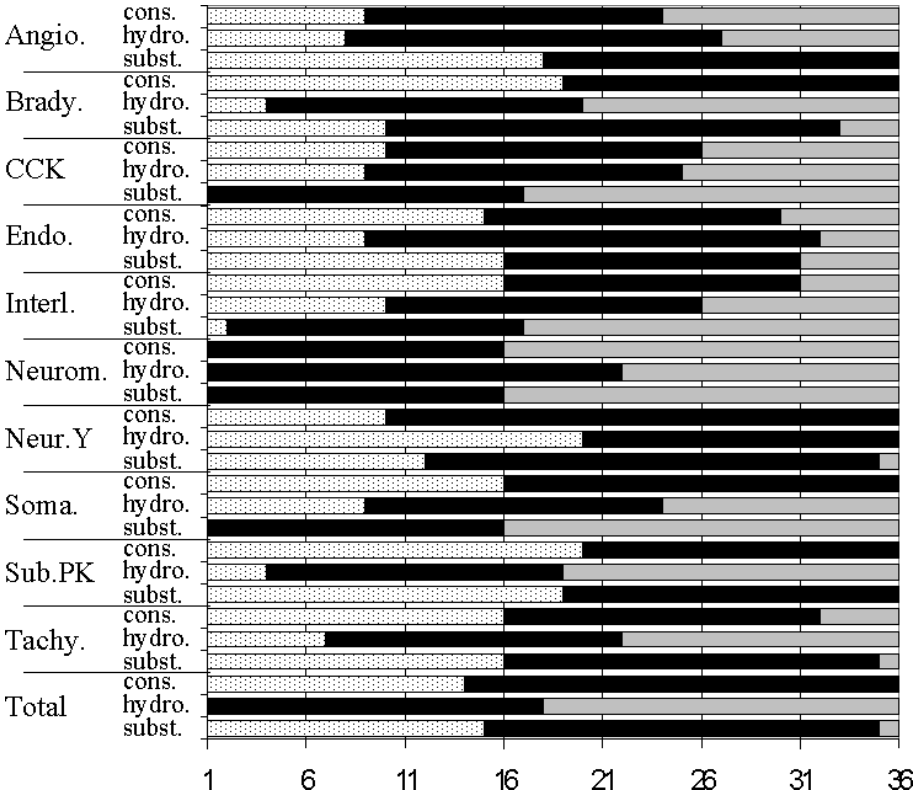


Figure 3.12: Transmembrane domains determined with the computer program Perscan for TMSS 5. The black columns are the predicted transmembrane domains and the light columns are situated at the extracellular side of the transmembrane domain and the dark grey columns at the intracellular side. For each alignment are three determination methods depicted in the order conservation, hydrophobicity and substitution method.

turn (RPT) value, which is calculated from the power spectrum and yields a value of 3.6 for an ideal α helix, was found near the ideal value for 80% of all alignments. A further parameter determined from the power spectrum is the α-helical periodicity index (AP). The AP gives a ratio of the extent of the periodicity in the helical region of the spectrum compared with that

over the whole spectrum. It is suggested that a value higher than 2.0 would indicate the presence of a region with significant helical periodicity. This is the case for 86% of the alignments, only 6% of the alignments show a value smaller than 1.6.
Figure 3.12 illustrates the TMDs which were calculated for the different alignments of TMSS 3 with three different determination methods of Perscan (hydrophobicity, conservation and substitution). It can clearly be seen that the TMDs calculated for the different alignments and determination methods are completely different, a finding that emerged for the analysis of all TMSSs.

3.3.10 TMD Determination with PredictProtein

By means of the PredictProtein Server [4] transmembrane helices were predicted via a system of neural networks. The multiple alignments of the peptide binding GPCRs of the seven TMSS were used as input data. The predicted TMDs for the seven TMSSs were found as: 8-30, 5-23, 11-28, 8-25, 10-29, 8-26 and 10-27, respectively. The reliability for all the seven predicted TMDs indicated by the note for reliability for the best model can reach a value of 9 on a scale of 0-9. All TMDs are longer than the TMDs obtained by Baldwin or obtained in this study, especially TMD 1 (23 residues). Except for TMD 2, all predicted TMDs from applying the neural network overlap almost completely with the TMDs predicted by Baldwin and by this study. It is remarkable, that the TMDs predicted by the neural network are extended towards the extracellular side of the membrane compared to the TMDs obtained by Baldwin. The prediction of the topology of all seven TMSSs is correct and is based on the difference between the positive charges at the extracellular and the intracellular side.

3.4 Discussion

The multiple alignments of the different families are considered as a first step for the identification of the potential TMSSs. Due to the low homology between the families in the loop regions a multiple alignment of only the putative TMSSs is beneficial. The fingerprint of Baldwin, which identifies the very conserved amino acids in the potential TMSSs, gives several anchor points for each multiple alignment of a TMSS. However, a precise determination of the boundaries of the TMDs is more reliable when it is based on a consensus approach comprising a variety of methods.
Studies carried out earlier by, e.g., von Heijne and Landolt-Marticorena have shown that the distribution of the amino acids in a TMD depends on its environment and on the membrane topology. The results obtained in this study confirmed their observations in great lines, only small disagreements were found for distribution profiles of, e.g., the residues valine and isoleucine. However, the membrane topology dependent distribution of charged residues is comparable with earlier studies. The distribution of tryptophan residues as a diagnostic membrane boundary residue is also observed by Landolt-Marticorena [98]. Woolf and Roux [113] claim that tryptophan is crucial for the orientation of gramicidin in the membrane and that the conformation of its sidechain fluctuates around well-defined orientations and is capable of forming hydrogen bonds with ester carbonyl groups of the lipids and with interfacial water. However,

[4] located at http://www.embl-heidelberg.de/predictprotein/predictprotein.html

this specific role of tryptophan could not be verified by Tieleman in molecular dynamics simulations on OmpF, Influenza M2, and alamethicin systems. Probably only a molecular dynamic simulation of a GPCR in a lipid environment will give detailed insights in this problem.
From this study it is clear that not all property profiles can be used for the determination of TMDs, since not all properties of the amino acids are membrane topology dependent. Especially the indices for bulkiness and α-helical conformation fail to display any correlation with the membrane topology.
The overall results stress the determining influence of the utilised scales and threshold values on the results, thus leading to mis-interpretations. Therefore, we preferred to follow a stringent approach by using normalised scores together with a threshold value of 0.2 for each TMSS.
The property profiles of the polarity scale of Zimmerman allow for a straightforward determination of the potential border residues of TMDs and turned out to be the profiles of choice, thus being superior to any hydrophobicity profile. In some cases, the property profiles show a clear α-helical periodicity (TMD 3 in Figure 3.2). The combination of the above-mentioned diagnostic profiles with the distribution profiles of characteristic residues such as Trp, Pro, Arg, Lys, Asp, and Glu clearly extends the reliability of the TMDBs determination.
The TMDs determined with the neural network system and predicted by the computer program Perscan deviate significantly from the TMDs determined by Baldwin and by the property and distribution profiles discussed here. Due to the large deviations in the TMDs from the different alignments and different determination methods of Perscan, the predicted TMDs are highly questionable.
It is generally accepted that the boundaries of the TMD cannot be assigned exactly, however they vary within approximately 2 amino acids at each side. This variation is not only due to the difficulties in determining the TMDBs by the applied methods, the property scales or the interpretation of the results. Also the dynamical structure of the membrane is an important factor for this uncertainty. The complexity of the dynamics of lipid molecules is a major reason for the membrane bilayer dynamics. The dynamics of lipids is not restricted to a lateral position, but they move also in the direction of the normal of the membrane. This mode of dynamics modulates the bilayer thickness and thus might influence the length of certain TMDs. Also the interfacial region between membrane and water is not a simple demarcation between a hydrophobic and hydrophilic environment, but it is considerably more complex than expected from simple macroscopic models.
A further structural factor, that is not considered in the applied methods, refers to the spatial extend of amino acid residues and the conformational flexibility of their sidechains. For example, arginine and lysine might be part of a TMD, while their long sidechains still reach the boundary region of the membrane, thus establishing interactions with the polar head groups of the membrane constituents. Indeed, these residues are capable of spanning two helical turns.
Further problems are imposed by the amphiphilic nature of entire helices, thus severely modulating the appearance of property and distribution profiles. For example, the property profiles of TMSS 2 (Figure 3.2) is strongly influenced by a highly conserved aspartate residue in the centre of the TMD giving rise to a significant distortion when compared to a typical TMD profile. Also the length of the property moment has influence on the variability in the angle of the property moment, e.g., a small vector can be strongly influenced by a single mutation.
A further aspect strongly influencing the property moments refers to the conformations of the transmembrane helices that are far from being ideal. Apart from kinks, these helices expose variable lipid accessible surfaces along their main helix axis. From the low resolution structure

of rhodopsin determined by Unger [62] it can be observed that helix 3 is more exposed to the lipid environment on its intracellular side, than on its intracellular part. This results in a dramatic change in size of property moments, thereby also modulating the angular orientation.
Helices 5 and 6 of the low resolution structure of bovine rhodopsin reveal significant kinks along their main axis, thereby rendering property moments for the entire helix questionable. A distinction of the kink in preceeding and following helical segments is therefore recommended. These factors influence not only the angular spread of the property moments for different alignments, but also influence the ideal anti-parallel conformation of the inside- and outside-pointing moments.
The results obtained from analysis of the experimentally determined protein structures (bacteriorhodopsin and the four-helix-bundle Felix) clearly showed that the property moments not always predict the inside of the corresponding helix bundle correctly. Most interestingly, not even the sequence and structural property moments were capable of predicting the TMDs consistently. Consequently, it is not sufficient to rationalise the low quality of predictions by non-ideal helix conformations.
The non-ideal conformation of the 3D structure of helices and the different surface exposure to the lipid environment seem to be important reasons for the discrepancy of the property moments at levels of +12 and -8 Å in (Figure 3.10) when compared to those computed for the 0 Å map. Figure 3.10 clearly reveals the interhelical space at the -8 Å level (intracellular) to be significantly smaller when compared to the +12 Å map (extracellular). This is underlined by the profiles of the bulkiness scale of Zimmerman where for the TMDs 2, 4, 5, 6 and 7 a significant lower score was observed for the intracellular side of the TMD.

3.5 Conclusion

From the above studies it turns out that automated methods for the prediction of TMDs fail to deliver reliable results, also because they are generally used as black box tools. Only a consensus approach comprising a combination of different methods based on a variety of parameter sets, which was applied on a series of different GPCR sequence alignments, yielded reliable results in terms of TMD determination. However, the TMD determination with a single residue resolution bears the danger of overinterpretation, since the dynamics of both, the protein as well as the membrane render a single residue resolution as a pseudo-accuracy.
Determination of TMDs with the discussed bioinformatic tools for the modelling of a GPCR are only a first step in this direction. The process of GPCR modelling is an iterative process comprising methods from bioinformatics, protein modelling, biochemistry, physics and further related scientific areas.

Chapter 4

Receptor-Modelling

A 3D model of the human CCK-B receptor based on the 3D structure of bovine Rhodopsin. A Molecular Modelling and Molecular Dynamics study.

"I believe the best test of a model is how well can the modeller answer the questions; what do you know now that you did not know before? And how can you find out if it is true."

- Jim Bower

econom

4.1 Introduction

Transduction of extracellular signals across the plasma membrane is generally achieved by the interaction of regulatory compounds with specific transmembrane receptor proteins. Approximately 80% of known hormones and neurotransmitters activate cellular signal transduction mechanism through binding to G protein-coupled receptors (GPCRs) [114], thus rendering members of this receptor superfamily as attractive targets for therapeutic intervention in a broad range of human diseases. While the biogenic monoamine-binding GPCRs are among the classical and well-established drug targets, the potential of the peptide- and protein-binding receptors from that superfamily is by far not yet exploited. Consequently, current medicinal chemistry programs pursued in pharmaceutical research are aimed to develop low-molecular weight, and non-peptide receptor agonists and/or antagonists respectively.

In this context, the development of active cholecystokinin (CCK) mimetics, selectively addressing the CCK-A or CCK-B receptor, has attracted the attention of biomedical research over the last decade [115, 69, 71]. The endogenous peptide ligand CCK exists in multiple forms of variable length, all generated from a 115-residue pre-prohormone, which possess a common C-terminus as shown, e.g., in CCK-8: Asp-Tyr(SO_3H)-Met-Gly-Trp-Met-Asn-Phe-NH_2. Precursor forms such as CCK-58, CCK-39, or CCK-8 are capable to bind to both CCK-receptors with nanomolecular affinities, while the non-sulphated analogues, the shorter penta- and tetrapeptideamides CCK-5, and CCK-4, as well as the related gastrointestinal hormone gastrin exhibit subnanomolecular affinities only towards the CCK-B receptor subtype [69, 71].

CCK-peptide derivatives used to be viewed predominantly as gut hormones responsible for the regulation of digestive processes with vague neurotransmitter activity in the central nervous system. This has drastically changed in recent years in that the CCK-A as well as the CCK-B receptors in the brain have emerged as mediators or modulators of anxiety, neuroleptic activity, arousal, analgesia and satiety [69]. Consequently, low-molecular weight compounds that selectively act on these receptors can serve as promising leads for the development of drugs for the aforementioned human disorders.

We have chosen the human CCK-B receptor as a working platform for proving the validity of a novel integrated molecular modelling procedure aimed to generate 3D structures of the receptor protein. The availability of 3D structure information on both interaction partners, that establish an activated ligand-receptor complex, will undoubtedly enrich our understanding on affinity, selectivity and receptor activation mechanisms.

Extensive biophysical, biochemical and immunological studies have revealed an enormous amount of structural data on these receptor proteins. These data suggest that the entire receptor superfamily show a common protein topology based on a membrane-spanning seven-helix bundle, generally termed the 7TM domain [6]. The amino terminus of the receptor protein is located within the extracellular compartment, while the carboxy-terminus resides in the cytoplasm. The transmembrane helices are sequentially connected by extra- and intracellular loops.

Apart from the X-ray structure of rhodopsin [5, 61] published in 2000, no experimentally determined high-resolution 3D structure of any GPCR, derived either by NMR-spectroscopy or X-ray crystallography, is available mainly due to the vagaries in purification and crystallisation of these membrane proteins. A low resolution 2D map of the electron density of bovine rhodopsin [57], a true GPCR accommodating the visual pigment retinal, provided first experimentally derived evidence for the existence of seven transmembrane helices. A low-resolution 3D structure of bovine rhodopsin [62] combined with projection structures of frog [58] and squid [59]

rhodopsin clearly confirmed the seven-helix bundle topology. These moderately resolved structures further demonstrated that although the membrane topology of bacteriorhodopsin and the GPCRs are similar, their 3D structures are significally different. The fact that bacteriorhodopsin is not a GPCR and displays only low sequence homology to any representative of this receptor family clearly disproves the idea that the X-ray structure of bacteriorhodopsin [42] could serve as a template structure for a molecular modelling driven attempt to generate GPCR models [116], even though this has been done for several years in numerous studies [117, 118, 119]. In those cases where bacteriorhodopsin was not used as structural core, the projection maps together with the low-resolution structures of rhodopsin were exploited, or models were constructed de novo [120, 121, 122].

Here we report on the application of a tailor-made modelling methodology especially suited for the generation of GPCR protein structures. The first and most critical step is the identification of the putative transmembrane sequence portions based on analysing multiple sequence alignments. In a previous study [123, Chapter 3], we performed a comparative evaluation of various protein sequence analysis tools for their capability to precisely predict the transmembrane domain boundaries. The study was conducted on the most prominent members of peptide-binding GPCRs, i.e., the human receptor sequences of angiotensin, bradykinin, CCK/gastrin, endothelin, interleukin-8, neuropeptide-Y, neuromedin, somatostatin, substance P & K and tachykinin receptors (ag22, ag2r, ag2s, brb1, brb2, cckr, gasr, et1r, etbr, il8a, il8b, ny1r, ny2r, ny4r, ny5r, nk3r, nmbr, ssr1, ssr2, ssr3, ssr4, ssr5, nk1r, nk2r, txkr). The different bioinformatics tools utilise algorithms tailor-made for the determination of potential transmembrane domains, such as periodicity analyses of different amino acid-associated properties, multiple sequence analyses, distribution analyses of amino acids, and directional helix descriptors of, e.g., hydrophobicity or conservation moments. In this comprehensive sequence analysis study [123, Chapter 3] we showed that the application of different methods leads to contradictory results in the prediction of transmembrane sequence stretches (TMSSs). However, by an especially elaborated combination of the results of the different methods, the putative TMSSs for a series of peptide-binding GPCRs could be determined reliably.

Based on these findings, we further expand our study on the question of a 3D structure model of the human CCK-B receptor. Special emphasis was laid on the computational refinement procedure applying specially developed simulation protocols explicitly accounting for the transmembrane environment of the receptor protein.

4.2 Methods

4.2.1 Bioinformatics

All Bioinformatics-based sequence analyses and molecular simulations were carried out on Silicon Graphics O2 workstations and a Silicon Graphics Origin parallel computer equipped with 20 R12000 processors. The overall strategy comprising the sequential application of different methodologies from bioinformatics, homology-related protein modelling and molecular simulation is summarised in Figure 4.1.

The amino acid sequences of the CCK-receptors of different species were retrieved from the protein sequence database SWISSPROT :

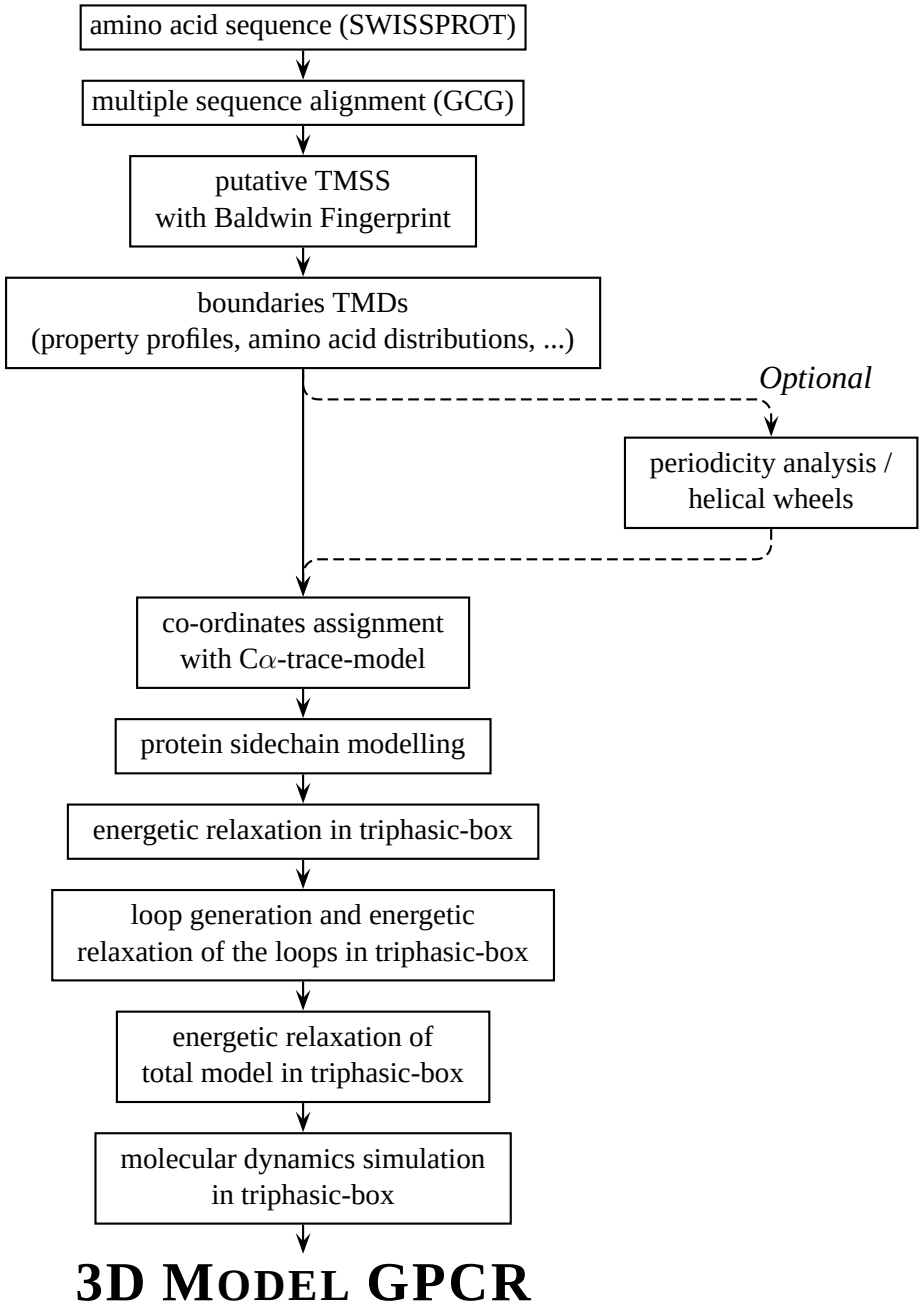


Figure 4.1: The general flow chart of the modelling of a 3D model of a G protein-coupled receptor.

entry-codes CCK-B receptors	species	ref.
gasr_prana	*Praomys natalensis*	[124]
gasr_rat	*Rattus norvegicus*	[125]
gasr_human	*Homo sapiens*	[126]
gasr_rabit	*Oryctolagus cuniculus*	[127]
gasr_canfa	*Canis familiaris*	[128]
entry-codes CCK-A receptors		
cckr_cavpo	*Cavia porcellus*	[129]
cckr_human	*Homo sapiens*	[130]
cckr_rat	*Rattus norvegicus*	[131]
cckr_xenla	*Xenopus laevis*	[132]

Multiple sequence alignments were obtained using the pile-up program as included in the GCG-package (Wisconsin Package Version 9.1, Genetics Computer Group, Madison wisc.; http://-www.gcg.com), considering the pam250 scoring matrix. The resulting sequence alignment was used to determine the putative TMSSs by superimposing the fingerprint-derived consensus patterns elaborated by Baldwin [60, 66].
The fine-tuning of the TMSS assignment of the human CCK-B receptor sequence was achieved following the procedure outlined in detail in ref [123].
Property profiles were computed for multiple sequence alignments based on a variety of amino acid-associated property scales, thus avoiding misleading conclusions that might arise from analyses based on one single amino acid sequence that eventually could be biased by single mutations. Profiles for four distinct properties were calculated: hydrophobicity according Kyte-Doolittle [78], hydrophobicity according to Sweet [91], polarity according to Zimmerman [92], and hydropathy according to Hopp-Woods [88]. The computations were performed on two multiple sequence alignments comprising: (i) 101 distinct sequences of peptide-binding GPCRs, and (ii) 9 sequences of CCK-receptors, respectively. For reasons of comparison the score of hydrophobicity profiles has been reversed and normalised, thus achieving that, e.g., regions of enhanced hydrophobicity exhibit a low score, similar to regions of reduced polarity. In addition to the property profiles, the distribution of particular amino acid residues [98, 133, 95] was used to differentiate between the transmembrane domains and membrane/water interface regions.

4.2.2 Molecular Modelling/Molecular Dynamics Simulations

Interactive model building and graphical analysis were done using the programs included in the molecular modelling software INSIGHT II (version 98.0) from Biosym Technologies (San Diego, CA). Molecular dynamics simulations were performed using the consistent valence force field without Morse potential or cross terms as implemented in the DISCOVER simulation package (version 2.98). According to the integrated modelling procedure, depicted in Figure 4.1, the transmembrane domain consisting of the seven helix bundle was generated based on the Cα-trace suggested by Baldwin [66] (co-ordinates were kindly provided by Joyce Baldwin). To account for favourable sidechain orientations, values for $\chi 1$ and $\chi 2$ of each helical residue were extracted from the rotamer library of Dunbrack and coworkers [134, 135, 136, 137]. Even though this library was derived from experimentally determined structures of globular, soluble proteins, it is assumed to provide reasonable starting conformations, due to the correlation of backbone conformation and sidechain orientation. The energetically most favourable rotamers

were assigned for $\chi 1$ and $\chi 2$ of each residue. In case of interatomic clashes between two sidechains of the same helix or two sidechains of adjacent helices, the second best rotameric state as assigned from the library to the affected amino acid sidechains.
After modelling the sidechain rotamers, a combined energy minimisation and molecular dynamics run was applied to energetically relax the structure of the transmembrane domain. The details of the simulation protocol employing an especially developed fixing and tethering scheme is given in Table 4.1. All simulations were carried out in the triphasic H_2O:CCl_4:H_2O solvent box [74, 75, 138] applying periodic boundary conditions. For this purpose, the receptor model was centred in the lipophilic core phase, with the helix-long axis aligned parallel with the Y-axis spanning across the bilayer. Consequently, both membrane-water interfaces, terminating the centre-membrane compartment, are parallel to the XZ-plane (Figure 4.1).
For the initial relaxations, the box dimensions were chosen to be 60x70x50 Å (X, Y, Z), whereby the thickness of the lipophilic phase was approximately 30 Å along the membrane normal (Y-axis). The remaining 20 Å along the Y-axis were soaked with water molecules on either sides.
A cut-off radius of 10 Å for calculating non-bonded interactions was used, while the neighbour list was updated every 10 fs. For the MD simulations, a time step for numerically integrating Newton's equation of motion of 1 fs was used with a weak temperature bath coupling.
In the initial phase of the simulation the helix backbone was fixed in order to prevent a substantial drift from the Baldwin template. The main emphasis of this simulation period is to remove conformational deficiencies from the crude starting conformation, while the overall seven-helix topology is allowed to relax in a later stage of the simulation protocol. Upon release of the backbone from fixation, distance restraints were introduced between O^i and NH^{i+4} to retain the internal hydrogen-bonding network of the helices. In contrast to the initial part, each helix is now allowed to perform translational and rotational reorientations within the context of the seven helix bundle.
After sufficient relaxation (Table 4.1) of the isolated transmembrane receptor protein domain, the extra- and intracellular loops were incorporated. In order to save computer time, the N-terminal 43 amino acids as well as residues 249 to 325 forming a large intercellular loop between helices 5 and 6 were not included into the model, since no conformational rationale for a meaningful co-ordinate assignment can be suggested. All included loops were constructed with an initially extended, unbiased backbone conformation avoiding intramolecular clashes with helix residues. For the subsequent simulations, the box dimension in the X- and Y-direction had to be expanded to 70x80x50 Å (X, Y, Z), whereby both aqueous phases occupy 25 Å and the central lipid phase 30 Å along the Y-axis.
A significant and experimentally confirmed [139, 140, 141] conformational restraint was introduced by forming a disulphide bridge between Cys-127 at the extracellular end of helix-3 and Cys-205 in the helix-4-helix-5 connecting loop, thus dramatically reducing conformational space principally accessible for this loop. For loop relaxation, the backbone atoms of all transmembrane residues were spatially fixed together with all sidechain atoms of the transmembrane core residues. To account for conformational consequences for interface residues due to the interactions with the loops, the sidechains of the terminal six transmembrane residues of each helix were allowed to move. Comparable to the simulation procedure of the transmembrane domain, all atoms were released for the final energetic relaxation using energy minimisation and molecular dynamics simulations. However, to retain the α-helical transmembrane structure, distance restraints were defined between the O^i and NH^{i+4} backbone atoms. The force constant scaling these hydrogen-bond restraints was gradually reduced in the course of the simulation.

Model	Calculation		Fixation	Tethering	Constraints
Calpha-trace-model modelling					
TMD-model	Minimise	SD & CG	Total protein	--	
	Templ. Force 1000 kcal/A	SD & CG	1,2,4,7 total, 3,6 BB	5 total, 3 & 6 side ch.	between 5 & 6
	Templ. Force 100 kcal/A	SD & CG	Idem	Idem	Idem
	Templ. Force 10 kcal/A	SD & CG	Idem	idem	Idem
	Minimise	SD & CG	Total protein	--	Idem
	Minimise	SD & CG	Backbone	--	Idem
	Minimise	SD & CG	Calpha-trace	--	idem
	Minimise	SD & CG	--	--	Idem
	Minimise	SD & CG	--	--	--
	Initialise Dynamics, 10 K	1000 steps	--	--	--
	Resume dynamics, 10 K	25000 steps	--	--	--
	Resume dynamics, 300 K	49000 steps	--	--	--
	Resume dynamics, 10 K	50000 steps	--	--	--
Modelling Extracellular Loops + N-terminus					
EL-TMD-model	Minimise	SD & CG	Total protein	--	BB helices, between 5 & 6
	Templ. Force 1000 kcal/A	SD & CG	Tmd aa's except teth. Atoms	Side ch. top 7 tmd-aa's	idem
	Templ. Force 100 kcal/A	SD & CG	Idem	idem	Idem
	Templ. Force 10 kcal/A	SD & CG	Idem	Idem	Idem
	Initialise Dynamics, 10 K	2000 steps	--	--	Idem
	Resume dynamics, 10 K	100.000 steps	--	--	Idem
	Resume dynamics, 150 K	100.000 steps	--	--	Idem
	Resume dynamics, 300 K	100.000 steps	--	--	Idem
	Resume dynamics, 150 K	250.000 steps	--	--	Idem
	Resume dynamics, 300 K	250.000 steps	--	--	Idem
Modelling 2 Intracellular Loops					
Total model	Minimise	SD & CG	Total protein	--	BB helices
	Templ. Force 1000 kcal/A	SD & CG	BB helices	Side ch. tmd-aa's	idem
	Templ. Force 100 kcal/A	SD & CG	Idem	Idem	Idem
	Templ. Force 10 kcal/A	SD & CG	Idem	Idem	Idem
	Minimise	SD & CG	--	--	Idem
	Initialise Dynamics, 10 K	2000 steps	--	--	Idem
	Resume dynamics, 10 K	100.000 steps	--	--	Idem
	Resume dynamics, 150 K	100.000 steps	--	--	Idem
	Resume dynamics, 300 K	500.000 steps	--	--	Idem
	Resume dynamics, 300 K	1.000.000 steps	--	--	--

BB = backbone
Side ch. = side chains
aa's = amino acids

SD = steep descent
CG = conjugate gradient

Templ. Force = Template force
Tmd = transmembrane domain

Table 4.1: The steps of the modelling procedure. The calculation method, the fixed atoms, the tethered atoms and the distance constains are listed for each calculation step.

To guarantee a sufficient consideration of the H-bonding network, a residual restraint was kept in our simulation protocol.
After the final minimisation, an extended molecular dynamics simulation covering a total of 1.15 ns was performed up to a target temperature of 300K. Snapshots were stored every 1.0 ps resulting in a trajectory comprising more than 1000 distinct structural frames.

4.3 Results and Discussion

4.3.1 Transmembrane Domain Determination

The multiple sequence alignment of nine CCK receptor sequences is depicted in Figure 4.2 (the Swissprot sequence identifier CCKR and GASR correspond to CCK-A and CCK-B receptor sequences, respectively). With respect to sequence identity, the CCK-receptor from the African clawed frog (CCKR_XENLA) cannot be clearly assigned to either the class of CCK-A or CCK-B receptors, since the identity ranges from 46% to 48% for any pairwise sequence comparison with CCKR_XENLA. Generally, the identity among sequences within a distinct subclass is found to be higher than 85%, while it drops to approximately 45% for sequence comparisons across the two subclasses [132].

Corresponding to its comparable sequence similarities with either CCK-A and CCK-B receptors, the *Xenopus laevis* CCK receptor displays indifferent binding characteristics to selective agonists and antagonists. Even though CCKR_XENLA seems to be a unique representative of CCK receptors, it was included in the subsequent procedure aimed at determining the transmembrane boundaries based on the multiple sequence alignment [132].

To identify the putative transmembrane sequence stretches of the CCK-B receptors, the fingerprint analysis based on the results elaborated by Baldwin [60, 66] was used. In Table 4.2 the alignment of the putative transmembrane sequence stretches of both, the human CCK-A and CCK-B receptor with the fingerprint signature reveals a significant mutual agreement, thus allowing for a preliminary determination of the transmembrane helical sequence portions.

The most commonly applied method to determine the transmembrane domains is the computation of a hydrophobicity profile based on a parameter set derived from the protein sequence under investigation. Apart from various hydrophobicity scales, several different amino acid-based property scales can be used to physicochemically characterise a protein sequence. A property profile is generated by a sliding window approach in that the desired property for a certain sequence position is computed as the mean value of that property averaged over a certain sequence window encompassing a certain number of flanking residues. For proteins with putative transmembrane sequence stretches, the choice of an appropriate window size is crucial [123, 79]. Most commonly, the window size is selected taking the entire span of the membrane bilayer and the insertion mode and angle of the transmembrane sequence into account. The minimal number of residues required to span a typical lipid bilayer is estimated to be 16 to 18 residues, when assuming a regular α-helical conformation and an orthogonal insertion mode. Consequently a window size of 17 is generally used to assign the transmembrane domains [79]. To achieve an enhanced resolution, especially at the transmembrane boundaries, a reduced window size of about one single helical turn (3 to 4 residues) was found to be more appropriate.

We computed property profiles solely based on the multiple sequence alignments of the putative TMSSs of the CCK receptors and additionally, based on the multiple sequence alignment of 101 GPCR sequences from ten different peptide binding GPCR families. The different profiles from different alignments and for different properties are depicted in Figure 4.3. By combination of selected property profiles with the distribution profiles of diagnostic amino acid residues such as Trp, Arg, Lys, His and Pro, the transmembrane domain boundaries were assigned. For example, according to the “positive-inside-rule” of von Heijne [95], Arg and Lys residues will be found more frequently in the cytoplasmic loops of a membrane protein, thus causing a non-

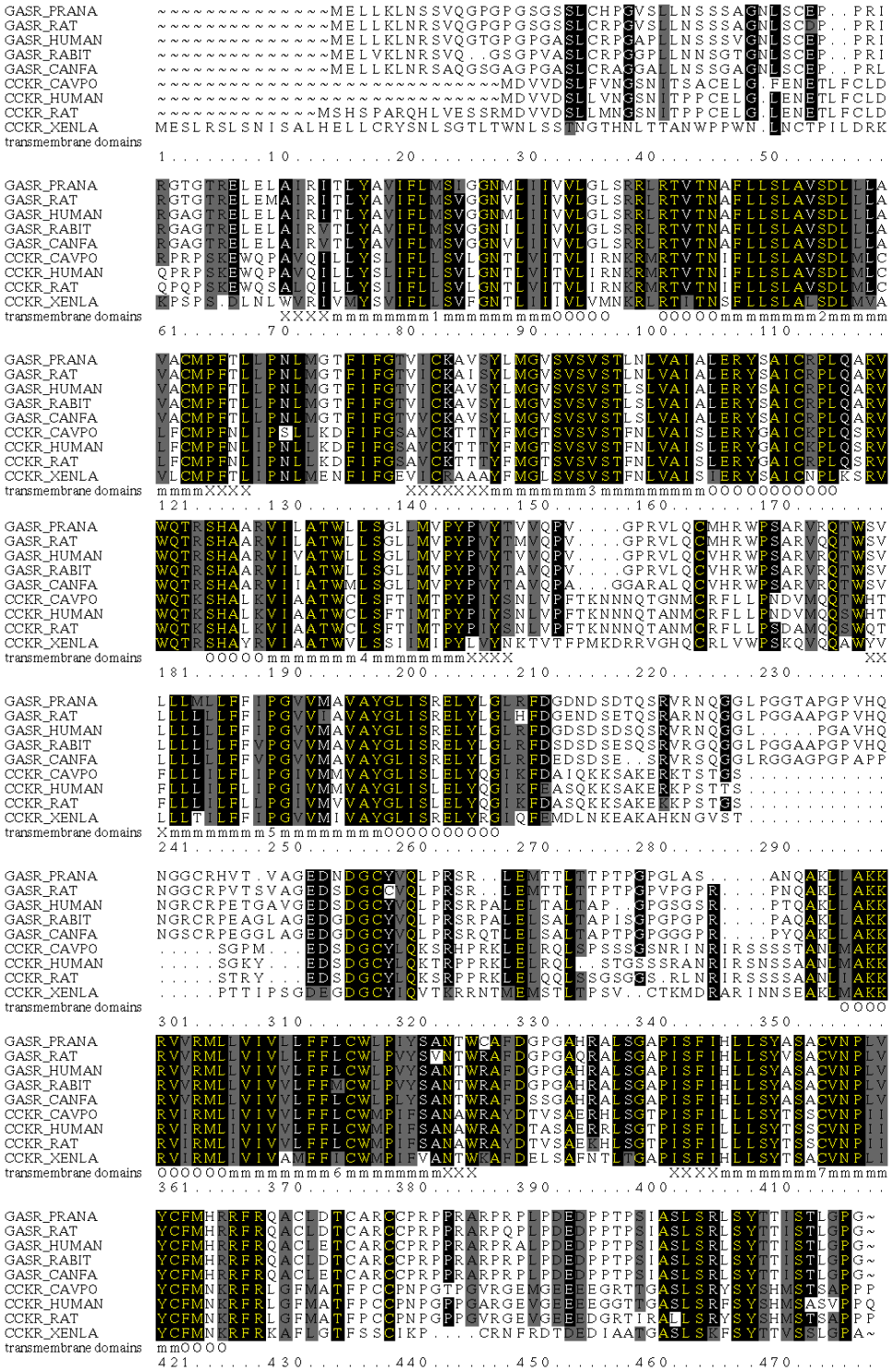


*Figure 4.2: Multiple sequence alignment of the nine sequences of the CCK-receptors (*black-box-yellow-letter *conserved residue in all CCK-sequences,* black-box-white-letter *identical residue in 80% of the CCK-sequences,* grey-box-black-letter *similar residue in 80% of the CCK-sequences, X = extracellular, O= intracellular and m = core segment of the helix of the Cα-trace model of Baldwin, inclusive number of the helix (alignment coloured with the program boxshade v3.31 of Kay Hofmann and Michael D. Baron; http://www.isrec.isb-sib.ch/software/Box-form.html)*

Helix I				Helix II				Helix III				Helix IV				Helix V				Helix VI				Helix VII			
numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint	numbers	cckr_human	gasr_human	fingerprint
1	K	R	#	36	L	M	#	1	G	G	+G	36	T	G	#	1	D	A	#					1	R	R	#
2	E	E	#	35	L	L	+	2	S	T	#	35	F	V	#	2	V	R	#					2	R	A	#
3	W	L	#	34	N	N	#	3	A	V	#	34	P	P	#	3	M	V	#	36	A	P	#	3	L	L	#
4	Q	E	#	33	P	P	#	4	V	I	+	33	V	Q	#	4	Q	R	#	35	T	G	#	4	S	S	#
5	P	L	#	32	I	L	+	5	C	C	+C(95)	32	L	V	#	5	Q	Q	#	34	D	D	#	5	G	G	#
6	A	A	#	31	L	L	+	6	K	K	#	31	N	V	#	6	S	T	+	33	Y	F	+	6	T	A	#
7	V	I		30	N	T	+	7	T	A	+	30	S	T	#	7	W	W	+	32	A	A	+	7	P	P	+
8	Q	R	+	29	F	F	+	8	T	V	#	29	Y	Y	+	8	H	S	#	31	R	R	#	8	I	I	#
9	I	I	+	28	P	P	+	9	T	S	#	28	I	V	+	9	T	V		30	W	W	+	9	S	S	#
10	L	T	+	27	M	M		10	Y	Y	+	27	P	P	+	10	F	L	+	29	A	T		10	F	F	+
11	L	L		26	C	C	#	11	F	L		26	Y	Y	#P(49)	11	L	L	#	28	N	N	#	11	I	I	+
12	Y	Y	+	25	F	A	+	12	M	M	#	25	P	P	P(68)	12	L	L	+	27	A	A	+	12	L	H	#
13	S	A		24	L	V		13	G	G	+	24	T	V		13	L	L		26	S	S		13	L	L	+
14	L	V		23	C	A	+	14	T	V		23	M	M	+	14	I	L		25	F	Y	#	14	L	L	L/F(83)
15	I	I		22	L	L	+	15	S	S	#	22	I	L		15	L	L	#	24	I	V	+Y/F(75)	15	S	S	
16	F	F	+	21	M	L	L	16	V	V	+	21	T	L		16	F	F	F(70)	23	P	P	P(100)	16	Y	Y	#
17	L	L		20	L	L	+L/F(80)	17	S	S	#	20	F	G		17	L	F		22	M	L		17	T	A	
18	L	M		19	**D**	**D**	#D(94)	18	V	V		19	S	S	S/A(79)	18	I	I		21	W	W	+W(85)	18	S	S	#N/S(80)
19	S	S	+	18	**S**	**S**	A/S(85)	19	S	S	S(77)	18	L	L		19	P	P	P(94)	20	C	C	C(71)	19	S	A	+S/C(76)
20	V	V		17	V	V		20	T	T		17	C	L		20	G	G	L	19	L	L		20	C	C	
21	L	G		16	**A**	**A**	A(86)	21	F	L	+	16	W	W	W(96)	21	I	V		18	F	F		21	V	V	+
22	**G**	**G**	+G(69)	15	**L**	**L**	L(96)	22	N	S	#	15	T	T		22	V	V		17	F	F	+F(82)	22	N	N	#N/D(99)
23	**N**	**N**	#N(99)	14	**S**	**S**	#N/S(88)	23	L	L	+L(74)	14	A	A		23	M	M	M/I(79)	16	L	L	+	23	P	P	P(98)
24	T	M		13	L	L	+	24	V	V		13	A	V		24	M	A		15	V	V		24	I	L	
25	L	L	L	12	L	L	I/L(77)	25	A	A		12	I	I	+	25	V	V	+	14	V	V		25	I	V	
26	**V**	I	V(76)	11	F	F		26	I	I	I(60)	11	V	V	+	26	A	A		13	I	I		26	Y	Y	Y(95)
27	I	I		10	I	A	+	27	S	A	S/A(87)	10	K	R	+	27	Y	Y	+Y(91)	12	V	V		27	C	C	+
28	T	V	+	9	N	N	#N(70)	28	L	L	+	9	L	A	#	28	G	G	+	11	I	L	+	28	F	F	+
29	V	V	+	8	T	T	#	29	E	E	#D/E(94)	8	A	A	+A	29	L	L	#	10	L	L		29	M	M	+
30	L	L	+	7	V	V	+	30	R	R	#R(99)	7	H	H	#	30	I	I	I/V(72)	9	M	M	+	30	N	H	#
31	I	G	#	6	T	T	#	31	Y	Y	+Y(79)	6	S	S	#	31	S	S	+	8	R	R	#K/R(79)	31	K	R	#
32	R	L	#	5	R	R	#	32	G	S	#	5	K	R	#	32	L	R	#	7	I	V	+	32	R	R	#
33	N	S	#	4	M	L	#	33	A	A	+	4	T	T	#	33	E	E	#	6	V	V	+	33	F	F	+F/Y(77)
34	K	R	#	3	R	R	#	34	I	I	I/V(94)	3	Q	Q	#	34	L	L		5	R	R	#K/R(75)	34	R	R	#R/K(79)
35	R	R	#	2	K	R	#	35	C	C	#	2	W	W	#	35	Y	Y	#	4	K	K	#	35	L	Q	#
36	M	L	#	1	N	S	#	36	K	R	#	1	V	V	#	36	Q	L	#	3	K	K	#	36	G	A	#
											P								#	2	A	A	#				
																				1	M	L	#				

Table 4.2: The fingerprint of Baldwin [60, 66] for the transmembrane domains of the rhodopsin-like GPCRs together with the sequences of the TMDs of the human CCK-A and CCK-B receptors.

symmetric charge distribution across the membrane, opposite to the charge distribution of the lipid molecules in the membrane bilayer [142, 143]. The putative TMSSs 4 and 6 of the CCK receptors and of the 101 peptide binding GPCRs clearly follow this "positive-inside-rule" of von Heijne, as can be seen in Figure 4.3.

The extracellular and cytoplasmic transmembrane domain boundaries are assigned based on tripeptide stretches exhibiting a steep decrease on the property profile followed by an extended sequence stretch with constantly low scores and a clear breaking point in the distribution profile. When comparing the property and distribution profiles based on the few CCK receptors (left panels in Figure 4.3), a less striking discrimination can be achieved. This finding can be attributed to the significant influence of a single mutation on the occurrence of the property and distribution profiles derived from the CCK receptor-base multiple alignments. The property profiles for TMD4 reveal a very flat non-conclusive profile for the CCK receptor-based alignment, while the property profile computed for the 101 GPCR sequence alignment displays the expected curve for a transmembrane sequence stretch. According to these results, we decided to use the profiles of the multiple alignments of the 101 peptide-binding GPCRs for determining the boundaries of the transmembrane domain.

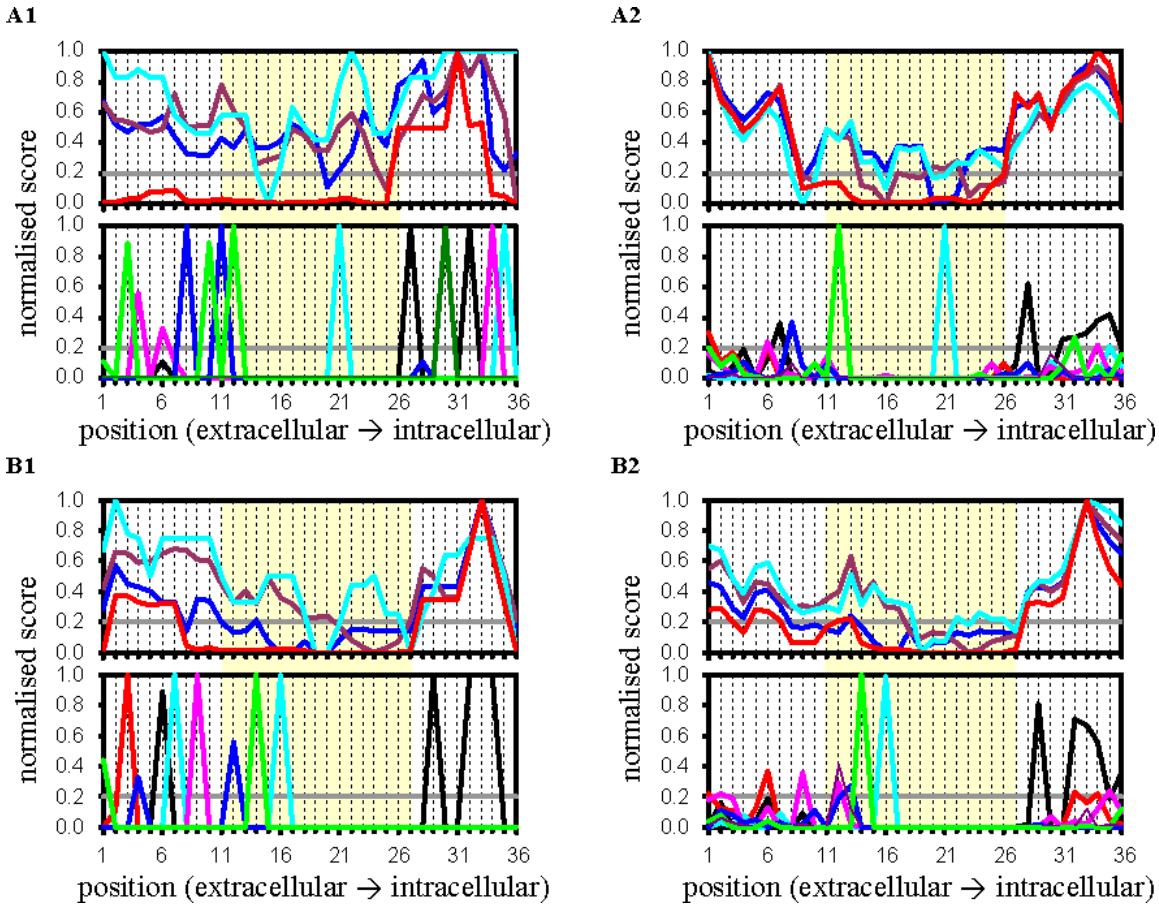


Figure 4.3: Property profiles and amino acid distributions of the multiple alignments of the Transmembrane Sequence Stretches 4 (A) and 6 (B) of the CCK-receptors (A1 & B1) and of the 101 sequences of peptide binding GPCRs (A2 & B2). In the upper part of each graph property profiles are shown with a window size of 3 (red: polarity scale of Zimmerman; blue: hydropathy scale of Hopp-Woods; light blue; hydrophobicity scale of Sweet; and purple: hydrophobicity scale of Kyte-Doolittle) and in the lower part amino acid distributions are shown (light blue: tryptophan; blue: tyrosine; red: aspatate & glutamate; green: proline; black: arginine & lysine; pink: asparagine & glutamine; and purple: histidine). Because these two TMSSs are even, the amino acid sequence numbering is in opposite direction in comparison to the position numbering (extracellular → intracellular = C-terminus ← N-terminus). The yellow strife is the determined transmembrane domain.

From a comparison of the different properties monitored over the sequence alignments, the property scale introduced by Zimmerman worked best in our hands. Boundary residues were assigned as amino acids flanking a stretch of at least 18 residues with scores smaller than 0.2 in the normalised Zimmerman polarity profile computed with a window size of 3. Within TMSSs 3, 6 and 7, residues are found in the transmembrane domain with score values between 0.2 and 0.3. Despite this slightly increased polarity score, these regions were assigned to be transmembrane domains since they are proceeded by sequence stretches with significantly smaller score values.

The transmembrane domain assignment obtained by analysing the fingerprint and property profiles was further refined by applying the following rules:

- Arg, Lys and Glu residues prefer to only occupy loop regions outside the transmembrane domains. However, if the TMD is too short to span the membrane, these residues might be found within the first helical turn (three residues), since their sidechains can extend beyond one helical turn to still place their charged sidechain functionalities into the aqueous compartment or near the polar headgroups of the lipids. [79]
- Due to their shorter sidechain, Asp residues can only occupy the terminal two positions of a transmembrane helix. [79]
- Trp residues are typical boundary amino acids and therefore indicate helix termination. [144, 98, 145]
- Non-highly conserved Pro residues can be found on the first three positions at the N-terminus and at the last position of the C-terminus of an α-helix, since the lost of a hydrogen-bond donor facility, disrupts the H-bond network within an helix. [98]

The final assignment of the transmembrane domain boundaries, together with the deviation from the Baldwin fingerprint approach, is listed in Table 4.3.

	TMDB determined by Baldwin					TMDB determined with window size 3					
	amino acid numbers of TMDB in TMSSs.		length of stretch	position of TMSS in graph †		determined with only prop. prof.		determined with prop. prof. + aa distr.		length of stretch	difference to TMDB determined by Baldwin
	extrac.	intrac.									
Helix 1	10	27	18	10	27	7	30	9	27	19	1	0
Helix 2	27	10	18	10	27	4	30	10	27	18	0	0
Helix 3	10	27	18	10	27	8	27	11*	27	17	-1	0
Helix 4	26*	11	16	11	26	9	26	10*	26*	17	1	0
Helix 5	11*	28	18	11	28	10	30	12	28	17	-1	0
Helix 6	27	11	17	10	26	8	27	10	27	18	0	1
Helix 7	12*	27	16	12	27	7	29	12*	28*	17	0	1

* = some difficulties with the determination of these transmembrane boundaries, see text.

† = the positions of the odd transmembrane sequence stretches are reversed numbered in the graphs.

Table 4.3: The numbers of the endings of the transmembrane domains of the helices deduced in a previous study of us. The used numbers are the numbers of table 4.2.

4.3.2 Molecular Modelling

The amino acid sequence stretches, identified as constituting the transmembrane domain, were mapped onto the Cα-trace model of Baldwin [66]. The amino acid sidechain torsions were assigned to the most favourable conformations extracted from the backbone-depending rotamer library for proteins developed by Dunbrack and coworkers [134, 135, 136, 137]. The resulting coordinate set for the helix bundle was subjected to an energy relaxation employing minimisation and dynamics simulation (for details refer to Table 4.1).

The initial model of the human CCK-B receptor revealed the termini of helices 5 and 6 to be separated too far for a compact helix bundle. Consequently, the simulation procedure was modified by introducing an elaborated distance constraint protocol attempting to bring the termini of helices 5 and 6 into close spatial proximity. Over these distinct template forcing and energy minimisations, helices 1, 2, 4 and 7 were kept fixed, while the sidechains atoms of helices 3 and 6 were allowed to move. A tethering scheme was applied to helix 5, in which the scaling force constant was gradually decreased from 1000 kcal/Å to 10 kcal/Å over the course of the minimisation.
Non-constrained standard minimisation concluded the refinement of the initial structure of the transmembrane domain of the human CCK-B receptor (for details refer to Table 4.1).

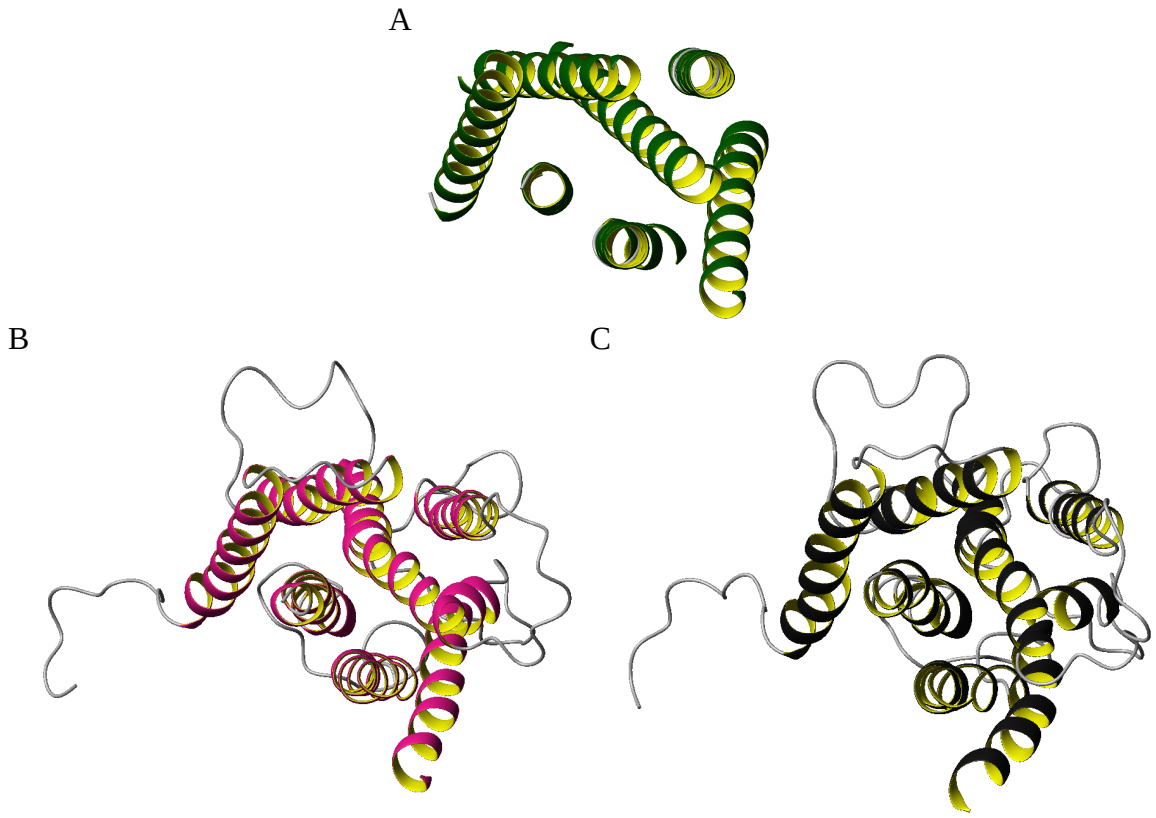


Figure 4.4: The Ribbon representation of the model from different modelling steps: (A) the start conformation with the co-ordinates of Baldwin for the Cα-trace, (B) the structure after the whole minimisation procedure, just before the final MD-run, and (C) the frame at 500 ps of the final MD-run. (The pictures are made with the program MolMol [146].)

The ribbon presentations shown in Figure 4.4 reveal the starting conformation that displays almost ideal α-helices. They are slightly distorted during the initial refinement process in favour of locally relaxed conformations and mutually attractive interhelical contacts.
After refinement of the seven-helix bundle, the extracellular and cytoplasmic loops were incorporated into the protein model by connecting the corresponding helical termini with the loop stretches in an expanded conformation. The modelled loops were minimised following the simulation protocol shown in Table 4.1.
In the initial part, the central core residues of the transmembrane helices were fixed, while the sidechains of the residues forming the terminal two helical turns (6 amino acid residues) were

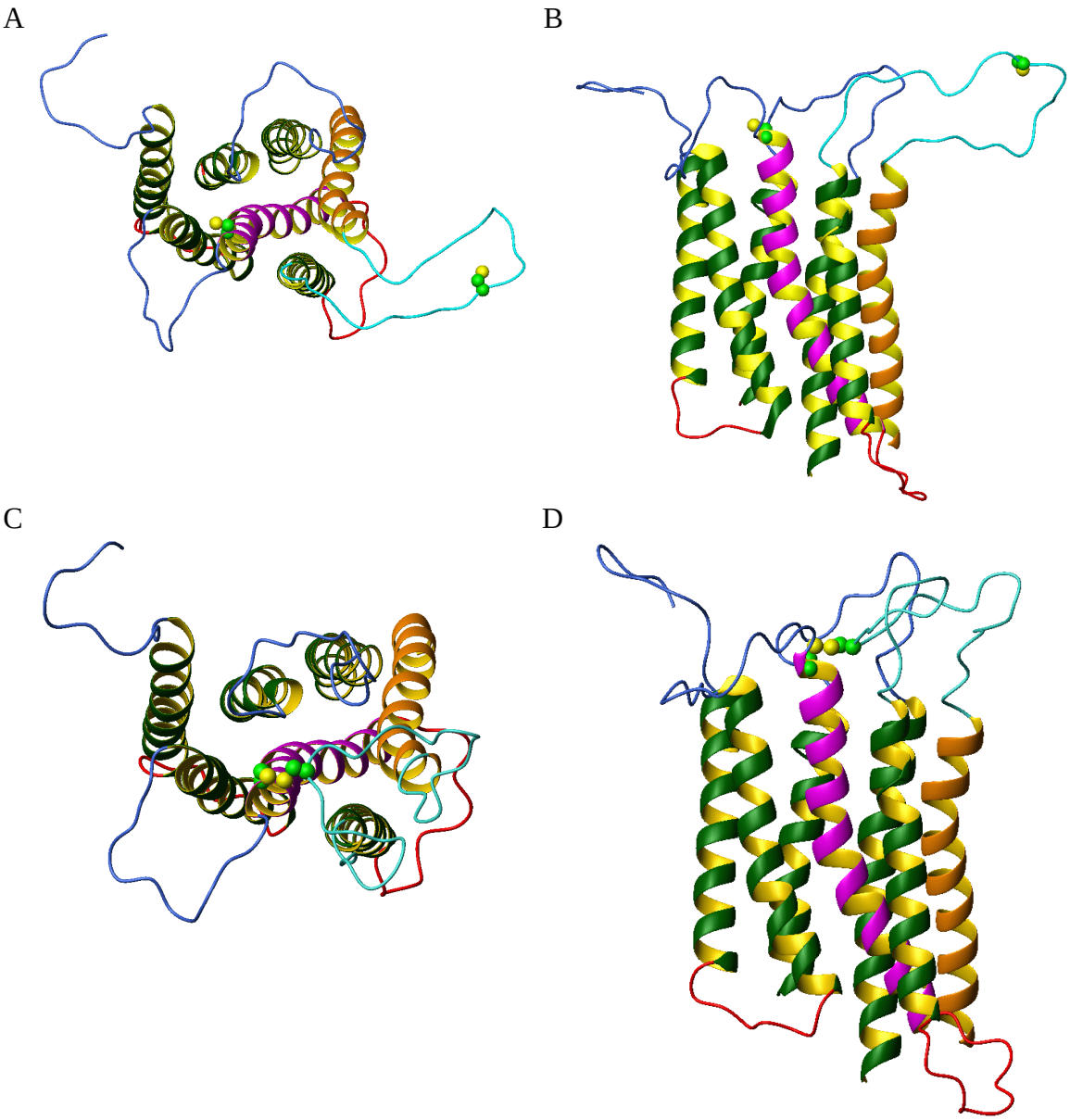


Figure 4.5: The topview (A) and the sideview (B) of the start conformation and the top view (C) and the side view (D) of the minimised conformation, respectively. Extracellular loops (blue), except extracellular loop 2 (cyan), intracellular loops (red), helix 3 (purple), helix 5 (orange) and two cysteine residues with CPK-model depicted. (The pictures are made with the program MolMol [146].)

free to move, since these can be estimated to be involved in interactions with loop residues. A distance constraint was introduced to allow for the formation of a disulfide bridge between the sidechains of Cys-127 at the extracellular end of helix 3 and Cys-205 within the second extracellular loop (see Figure 4.5). The distance between the two involved sulfur atoms of the bridging Cys residues changes from initially 40 Å to approximately 2 Å. In the final part of that relaxation strategy, the entire model of the receptor protein was energetically minimised.
After energy minimisation, an annealing strategy, employing MD, was chosen composed of a tailored temperature profile characterised by a heating and cooling procedure (Table 4.1). To allow for translational and rotational freedom of each transmembrane helix, no fixing or positional restraining was imposed on the protein model in the final refinement simulation. Distance constraints between backbone carboxyl oxygen atoms (hydrogen bonds acceptors) and amide groups (H-bond donors) were applied in a $O^{i} \longleftrightarrow$ H-N^{i+4} relation to conserve the helical conformation of the transmembrane helices.
After sufficient relaxations, the entire system was subjected to a 1.15 ns molecular dynamics run that is used for structure analysis. The stability of the receptor protein model is monitored over the simulation by following the root-mean-square deviation for structural superposition of the transmembrane domain residues of distinct conformations onto the starting conformation (Figure 4.6). Apart from a significant deviation in the initial part of the refinement procedure,

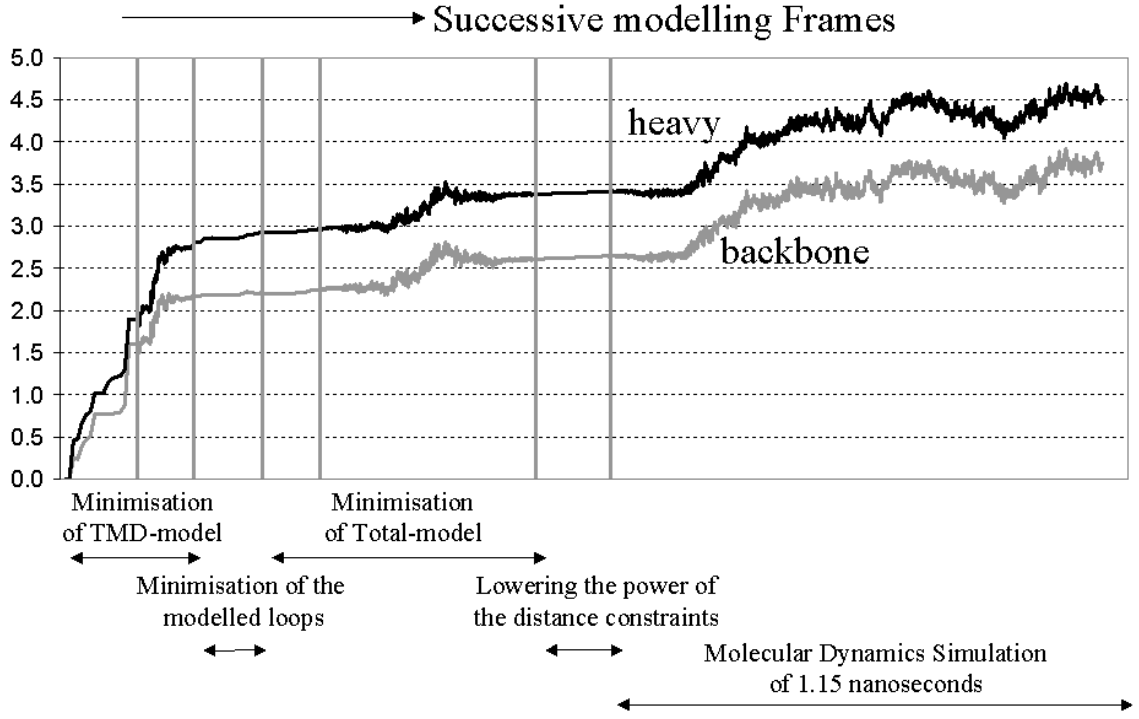


Figure 4.6: The Root Mean Square (RMS) deviation of the backbone (black) and heavy atoms (grey) of the TMDs during the total modelling procedure. The start conformation is taken as the reference structure. The frames of three MD sections, the second part of the minimisation of the TMD-model, the second part of the minimisation of the total model and the final molecular dynamics simulation, were shrinked (spacing is 10% of that of the other frames) for a better view.

the protein structure exhibits a reasonable stability as indicated by an rms deviation of around

1 Å within the final MD simulation conducted over 1.15 ns after initialisation. This finding is corroborated by the energy diagrams monitored over the production period of the long-term MD simulation (Figure 4.7). After the initial 300 ps of initialising, the entire system reaches steady energy and temperature levels without any remarkable fluctuations within the remaining simulation period.
Consequently, the final 750 ps (frames 301 to 1050) were used for analysis.

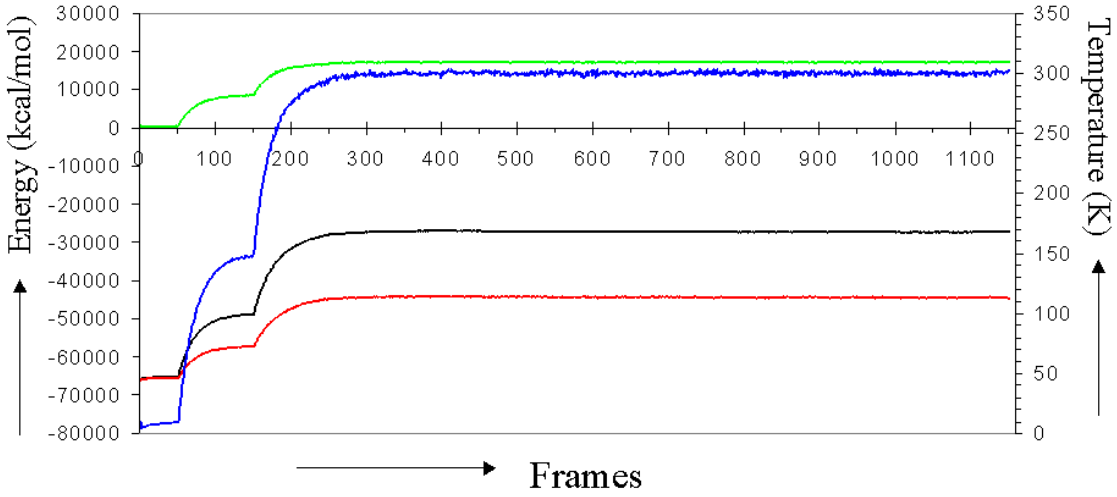


Figure 4.7: The total (black), kinetic (green) and potential (red) energy of the complete system during the final molecular dynamics simulation and the temperature (blue) during this MD-run (right Y-axis).

4.3.3 Production Period of the Molecular Dynamics Simulation

The final 750 ps of the 1.15 ns MD simulation were used to analyse the conformational stability of the generated receptor protein model. For assessment of the vertical alignment of the receptor protein's location within the tri-phasic membrane-type environment, the standard deviations of the Y-coordinate component of selected atoms were computed, Y being oriented perpendicularly to the membrane surface plane (XZ). Large deviations of especially the backbone atoms of the CCK-B receptor core domain would indicate a migrational component of the protein during the MD simulation, thus revealing an incompatibility of the protein model with its non-isotropic environment. The analysis of 750 snapshots stored in intervals of 1 ps uncovers only minor standard deviations for the Cα atoms of the entire protein model (Figure 4.8A) with respect to the Y-coordinates. From the same analysis applied to sidechain atoms, moderately increased flexibility in the $\pm$Y direction is observed for the transmembrane residues, while flexibility is increased for loop residues (Figure 4.8B). The entire analysis depicted in Figure 4.8 clearly reveals that the protein model perfectly aligns with its location in the tri-phasic solvent system, since no overall migrational component along the Y-axis can be observed. This provides additional confidence on the overall compatibility of the different protein domains (extracellular loops / transmembrane helix bundle / cytoplasmic loops) with the exposed environment.
The distribution pattern of the protein flexibility profile monitored over the final 750 ps of the MD simulation is illustrated in the ribbon model shown in Figure 4.9, in which the thickness

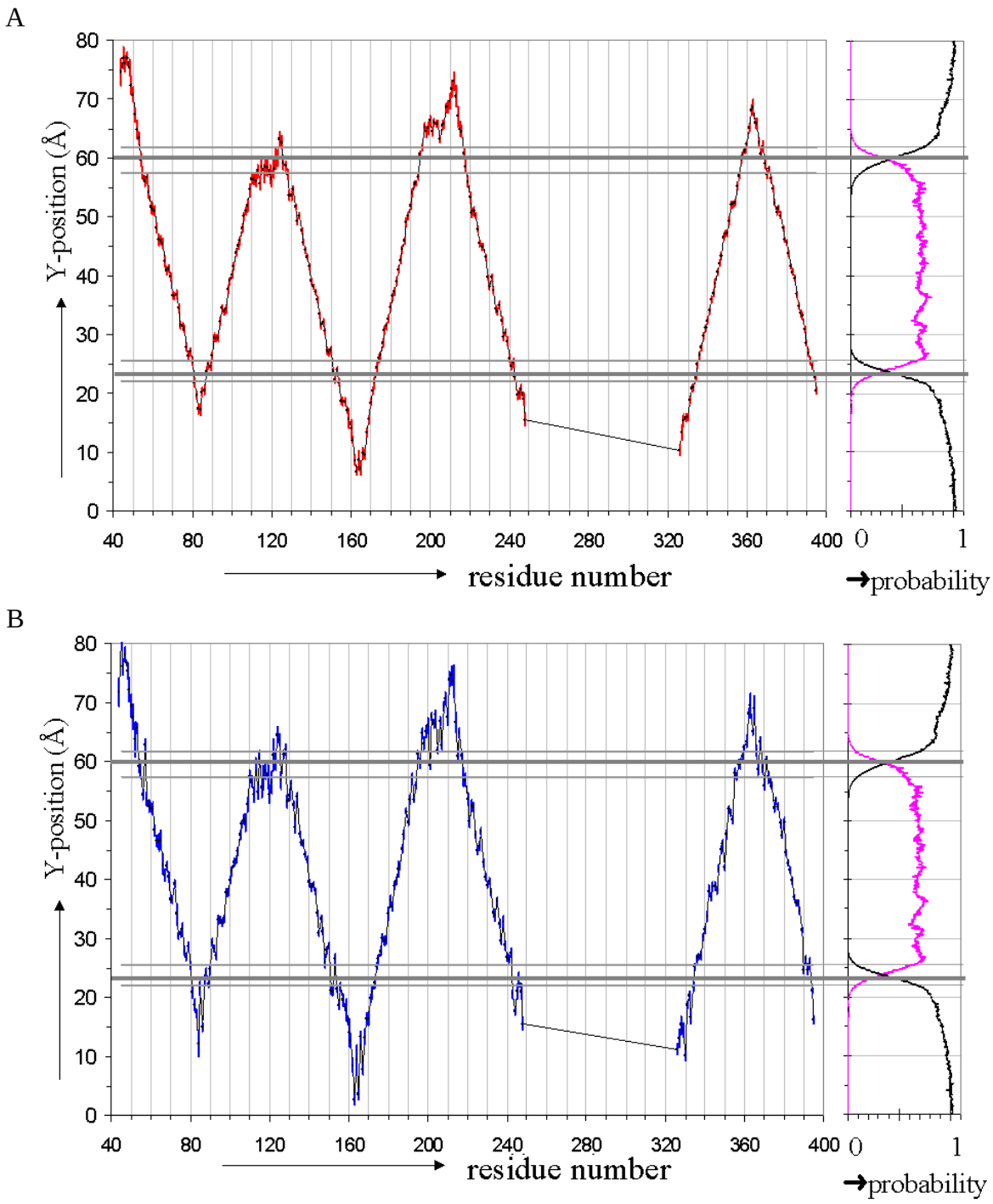


Figure 4.8: The average Y-positions of the Cα-atoms (A) and one atom of the sidechain (B) of each residue during the 750 ps of the production period of the final MD-run. The length of the coloured bar on each point represents the standard deviation in the Y-position of the atoms. On the right hand the probabilities of the water (black) and the CCl_4-molecules (magenta) are depicted. The Y-postions of the boundary planes (dark grey) and the interfacial regions (grey) are depicted with horizontal lines.

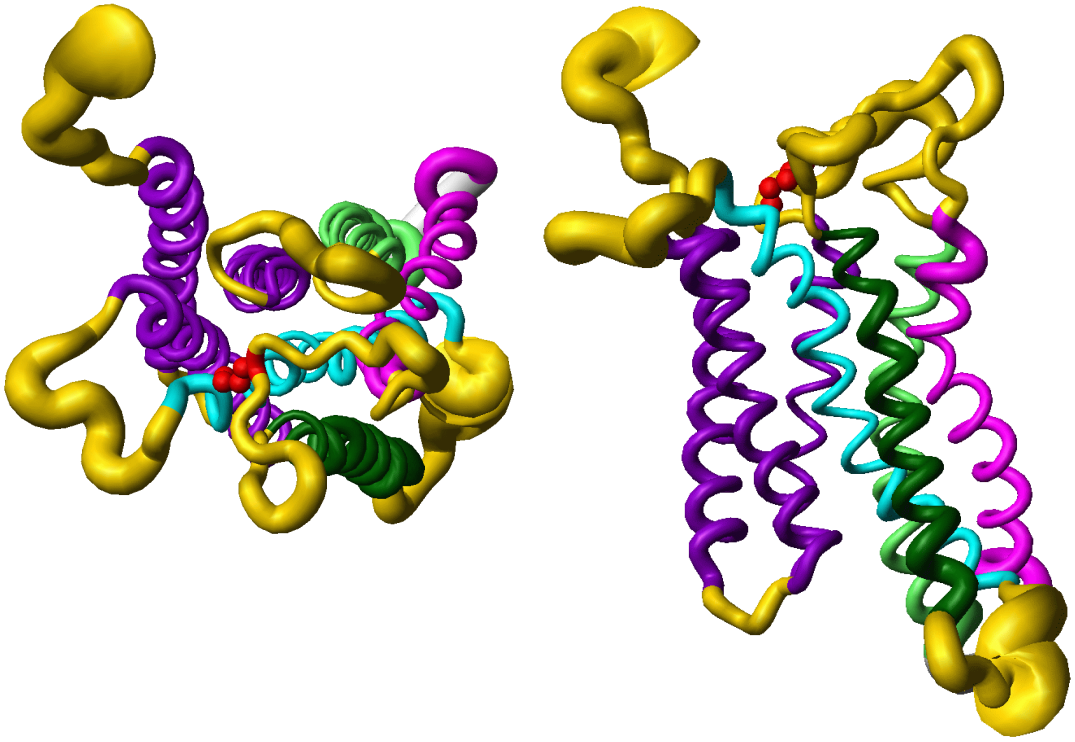


Figure 4.9: "Sausage" representation of the flexibility of the backbone atoms of the total model during the production period of the final molecular dynamics simulation (Cys-bridge = red, helix 3 = cyan, helix 4 = dark green, helix 5 = magenta, helix 6 = light green, helices 1, 2 and 7 = dark violet, loops = yellow, and 'cutted intracellular loop 3' = white, the pictures are made with the program MolMol [146].)

of the backbone ribbon correlates with the computed root mean square deviations from the average structure for all Cα atoms. Not surprisingly, the N-terminus which is exposed to the extracellular aqueous compartment exhibits highest flexibility. The first and third extracellular loops together with the second intracellular loop display remarkable conformational flexibility as well, while the second extracellular loop is clearly constrained by the explicitly depicted disulfide bond. The 7TM helical domain demonstrates a remarkably reduced flexibility with helix 3 being most restrained (Figure 4.9).

The overall stability of the receptor protein, monitored over the MD simulation within its native-like environment, is highlighted in this context by the observation that even distinct water molecules emerge as integral components of the receptor protein structure, as exemplified in Figure 4.10. The depicted water molecule is engaged in a stable hydrogen-bond network, thus relaying hydrogen-bonds between residues Cys-107, Thr-111 (sidechain OH), and His-376 (sidechain imidazole), respectively. This particular water molecule is found to be engaged in this interaction network within all 750 snapshots covering the final production part of the MD simulation.

Further details of the analysis of the final 750 ps of the MD simulation are depicted in the Tables 4.4 - 4.7. Available structure-relevant information as the fingerprint of Baldwin [60, 66] (row 3), general structure information of GPCRs (e.g. positive-inside rule data; row 4) and site-directed mutagenesis data of both CCK-receptors [147, 148, 149, 150, 151, 152, 153, 154] (row 5) are aligned with structural dynamical properties as hydrogen bond occurrences (row 8

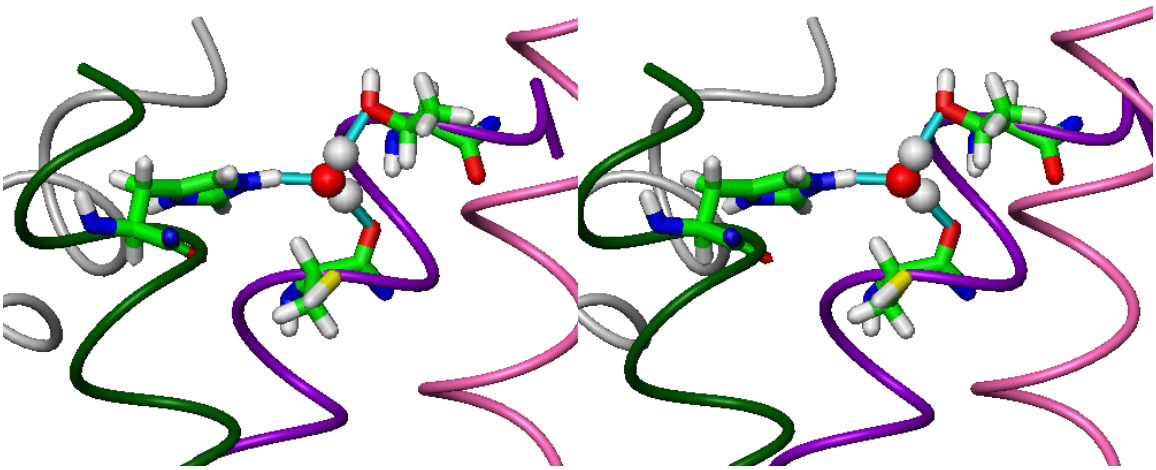


Figure 4.10: Side-by-side stereoview of a structural detail including a water molecule which is strongly bound between the residues Cys107, Thr111 and His376 and remains fixed during the whole 750 ps production period of the final molecular dynamics simulation (helix 1 = magenta, helix 2 = dark violet, helix 3 = white and helix 7 = dark green and the hydrogen bonds are represented by cyan coloured bonds, the picture was made with the program MolMol [146]).

and 9) and spherical distribution functions with a radius of 5 Å of the solvent (row 6) and of the protein atoms (row 7) around both, backbone and sidechain atoms of all the modelled amino acid residues. It can clearly be seen that the spherical distribution function of both, the solvent and the protein atoms follow the fingerprint of Baldwin almost perfectly, indicating that the model of the human CCK-B receptor is in good agreement with this fingerprint. Furthermore, the mutated transmembrane domain amino acids with no effect on ligand binding are almost all pointing to the membrane environment and the amino acids within the transmembrane domains showing any effects on ligand binding are pointing towards the protein interior. The $O^{i} \longleftrightarrow$ H-N^{i+4} hydrogen bond network to conserve the helical conformation of the transmembrane helices is retained throughout the entire MD simulation and also several interhelical hydrogen bonds were found within more than 80% of all 750 snapshots covering the final production part of the MD simulation.

Summarising, the modelling-derived protein structure of the human CCK-B receptor, described above, remains stable over a 1.15 ns MD simulation and displays a conformational flexibility pattern that is expected for a transmembrane protein.

Furthermore, the modelled protein structure of the human CCK-B receptor was compared with the crystal structure of bovine rhodopsin [5]. The root mean square deviations between the model of the receptor and the X-ray structure were comparable with those found by Bourne and Meng. Bourne and Meng [61] described the superposition between the crystal structure of rhodopsin and three recent model structures and found that the Cα-trace of the transmembrane helices fitted reasonably well (RMS-deviations of 3.1 - 3.2 Å for the Cα-trace of the seven transmembrane helices). The root mean square deviation between the X-ray structure (PDB-code 1F88 [5]) and the model structure of this study was for the backbone atoms (Cα, C and N) of the complete α-helices 4.5 Å and for the transmembrane domains 3.9 Å, respectively. The two α-helices with the highest RMS-deviations in their transmembrane domains were α-helices 2 and 5, namely 2.3 and 2.6 Å, respectively. Helix 6 is a rather straight α-helix and has only a RMS-deviation of 0.9 Å for its backbone atoms within the transmembrane domain.

4.4 Conclusion

Given the conformational characteristics described above that were obtained from a sophisticated molecular modelling and simulation procedure developed in this study, we feel confident to use the derived protein model for further studies that are aimed to shed light on the structural details that govern ligand binding.

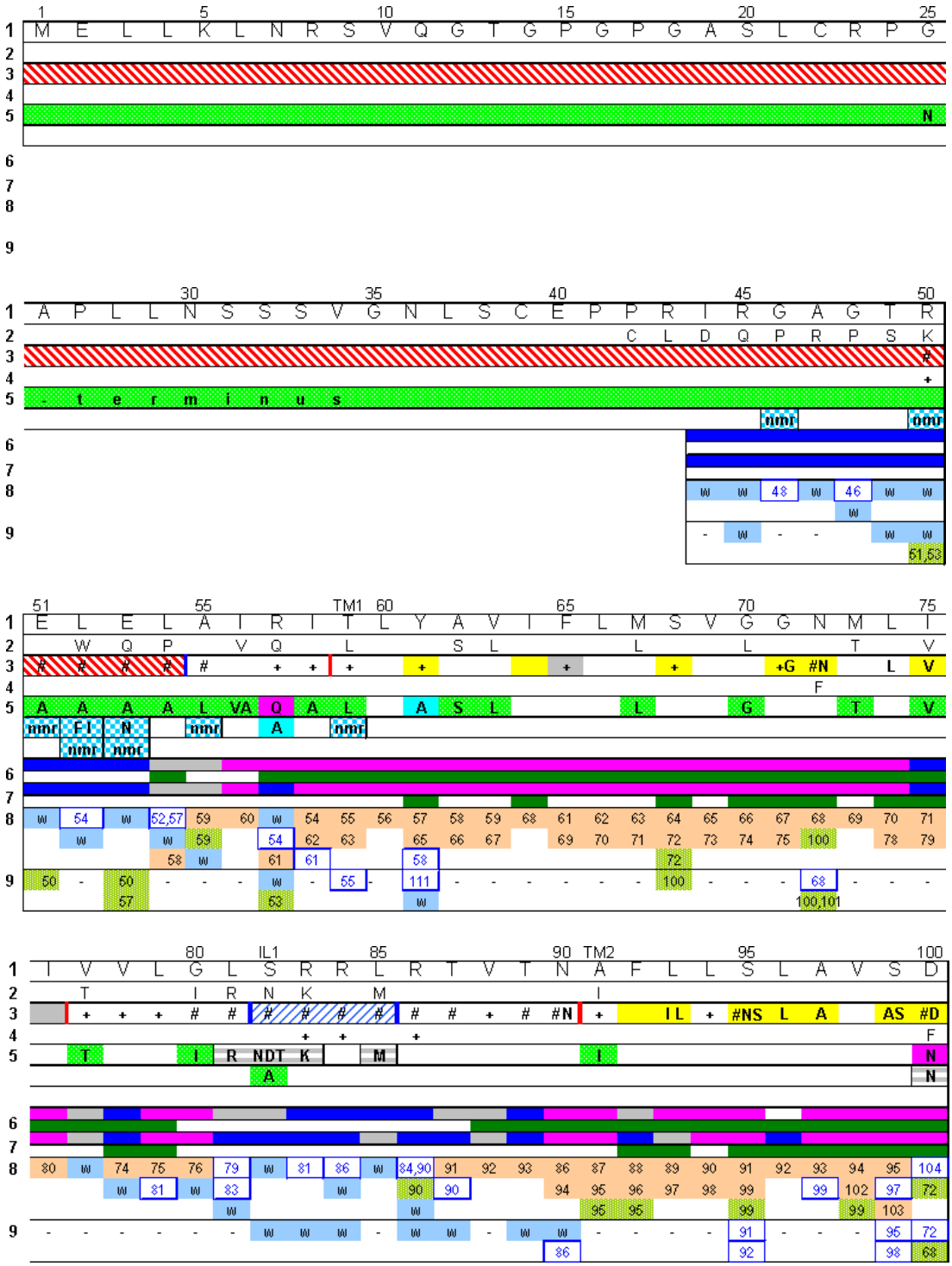


Table 4.4: The amino acid sequence of the human CCK-B receptor PART I, colour-coding see Table 4.7.

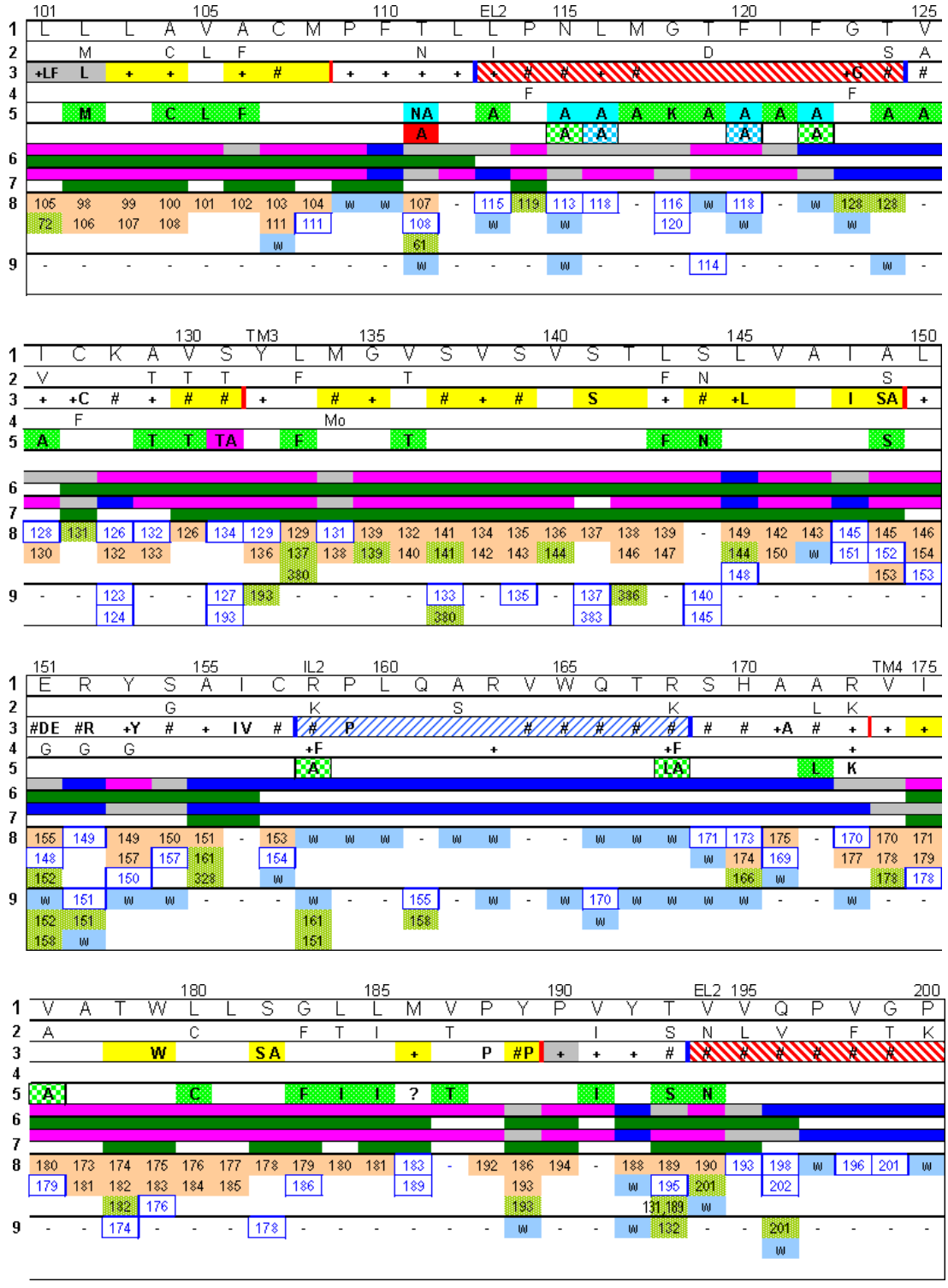


Table 4.5: The amino acid sequence of the human CCK-B receptor PART II, colour-coding see Table 4.7.

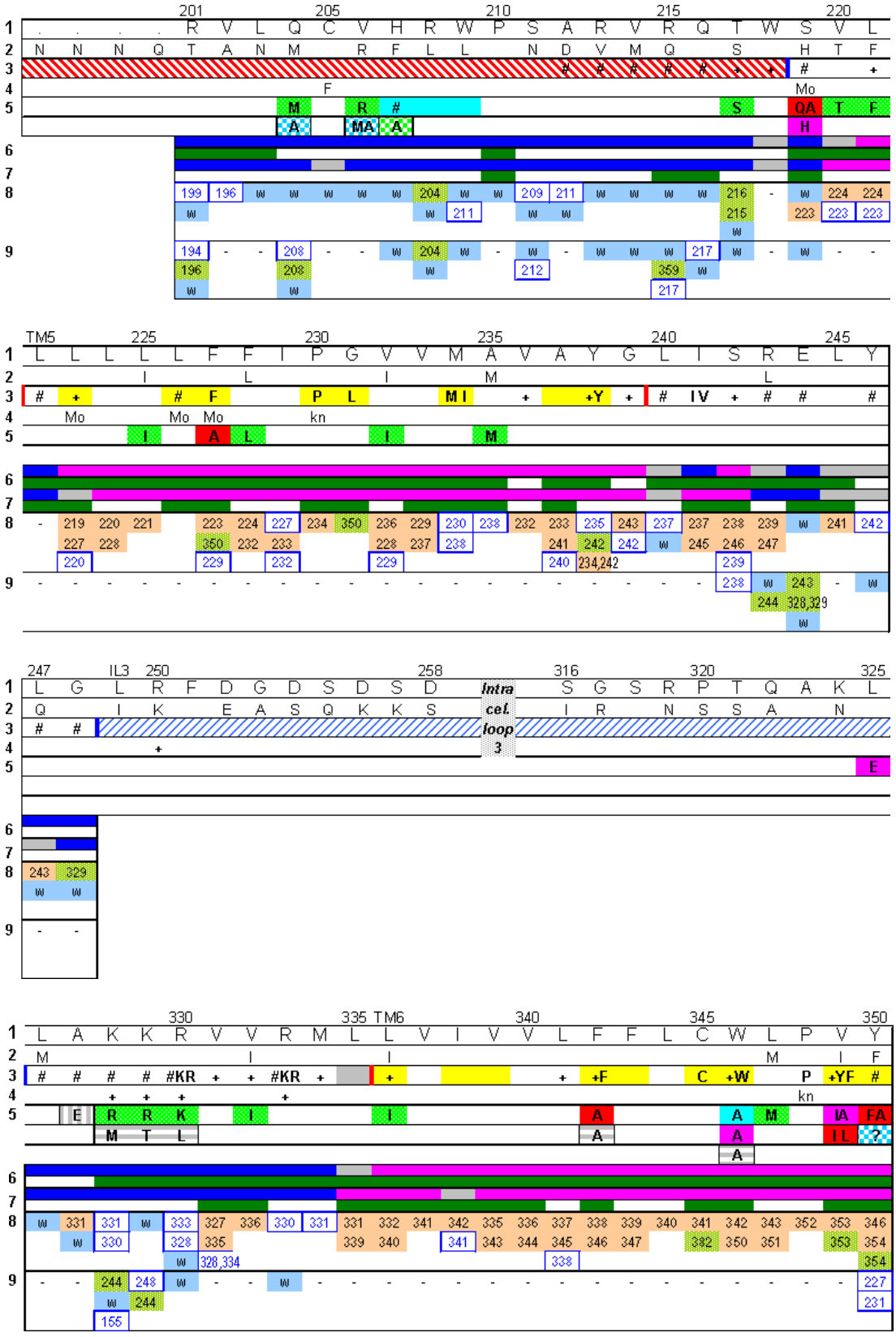
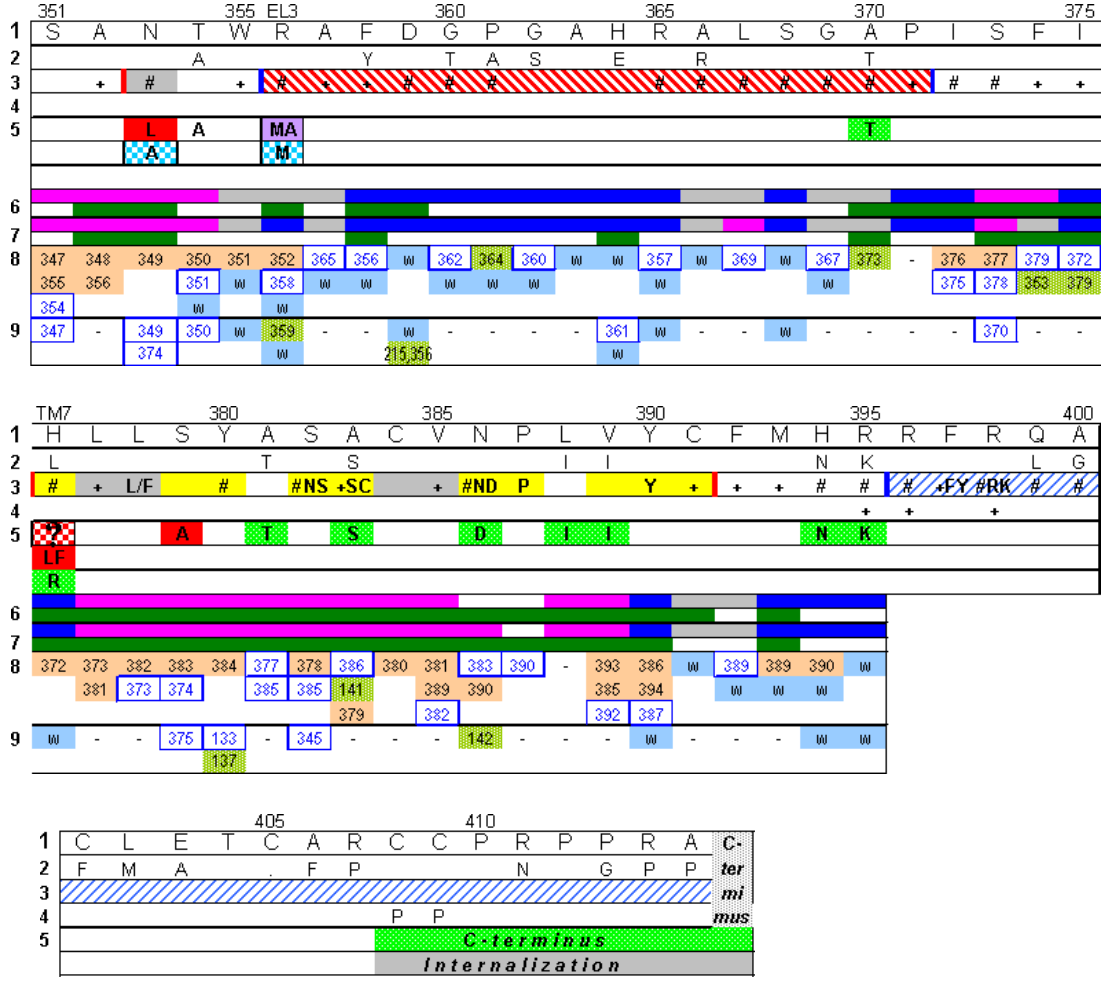


Table 4.6: The amino acid sequence of the human CCK-B receptor PART III, colour-coding see Table 4.7.



row 1	amino acid sequence of the human CCK-B receptor
row 2	different amino acids of the human CCK-A receptor compared to the human CCK-B receptor
row 3	Fingerprint of Baldwin; red = extracellular region, blue = intracellular region, yellow = inside protein, grey = on border of the protein, x = conserved amino acid, + = polar residue in a few sequences, # = polar residue in >10% of the sequences
row 4	general concepts; F = important for folding, G = G-protein binding, mo = binding ligand monoamine GPCRs, + = positive-inside rule, kn = knick in helix
row 5	known mutations of the CCK-B receptor: red = affect antagonist binding, blue = affect agonist binding, magenta = affect agonist and antagonist binding, green = no affect on binding agonist or antagonist or signal transduction, horizontal grey strifes = affect signal transduction; same effects but than for the CCK-A receptor: spickeled red, blue, green, dunkle magenta and vertical grey strifes.
row 6	spherical distribution function around Cα-atom of each residue (5Å); first part: blue = >70% water, >70% magenta = CCl_4 and grey = mixed; second part: green = inside protein, white = outside protein
row 7	spherical distribution function around specific sidechain atom; colours idem to row 6
row 8	H-bonds excisting in >80% of the 750ps for the main atoms; w-blue-box = H-bond with water, blue-residue-number = with backbone atoms of that residue, residue-number-in-orange-box = α-helical H-bond with that residue, residue-number-in-green/yellow-box = H-bond with sidechain of that residue.
row 9	H-bonds excisting in >80% of the 750ps for the sidechain atoms; colours idem to row 8

Table 4.7: The amino acid sequence of the human CCK-B receptor PART IV.

Chapter 5

CCK Ligands

Overview on CCK agonists and antagonists.

"Words are, of course, the most powerful drug used by mankind."

- Rudyard Kipling (**1865-1936**), poet/writer and Nobel Prize in Literature laureate

5.1 Introduction

As mentioned in section 2.2, CCK-8 is the predominant form of CCK (cholecystokinin) which displays biological activities, both in the peripheral and central nervous system. These different areas are responsible for the notation of the two different CCK receptors, notably CCK-A (alimentary) and CCK-B (brain), even though it has been indicated that CCK-B also occurs in the periphery and CCK-A in the brain, respectively.
The broad therapeutic potential of CCK-ligands (Figure 5.1) resulted in considerable research efforts. The complexity of CCK-receptor modulation in different physiological processes renders the exact prediction of its physiological and therapeutic interference somewhat difficult, accordingly the role of agonism as well as antagonism of CCK-A or CCK-B receptors is not entirely understood.
Potent and selective peptide and non-peptide agonists and antagonists have been discovered for both types of CCK-receptors. It has become obvious in recent years that therapeutic application of CCK-receptor antagonists provides a broader spectrum of therapies than the use of agonists. However, this can be partly attributed to the fact that the nonpeptide CCK agonists were discovered only recently. The discovery of the benzodiazepine containing asperlicin in 1985 [115], a natural product isolated from *fungus*, marks a milestone in the design of CCK-ligands. Based on this structure, several very potent and selective compounds have been synthesised. The prototype antagonist for the CCK-A receptor subtype is L-364,718 (devazepide). For the CCK-B receptor subtype L-365,260 emerged from asperlicin-based medicinal chemistry projects (Figure 5.2).
Over the past decade, a large number of highly selective agonists and antagonists acting selectively on CCK-A or CCK-B receptors have been developed (reviewed in [155, 29, 156, 157, 158]). This chapter will give a brief overview on the principally devised types of lead compounds. A thorough and exhaustive discussion of CCK-A and CCK-B receptor mediating compounds is clearly beyond the scope of this thesis. Special interest will be focused on those compounds acting on the CCK-B receptor which is the receptor of our predominant interest.

5.2 CCK-A Agonists

For the design of CCK-A agonists the CCK peptide was used as lead compound. Since it was known that the C-terminal portion of the peptide is mainly responsible for receptor binding, systematic variation of chain length together with the concept of conformational restriction revealed compounds with enhanced selectivity and considerable proteolytic stability.
A research group at Hoffmann-La Roche changed different positions of the acylated heptapeptide CCK-7, e.g., Phe into MePhe, Met into Nle and Asp into Thr(SO_3). Two compounds, Ro 23-2154 (**3**) [1] and Ro 23-7014 (**4**) [161] (Figure 5.3), displayed improved selectivity for the CCK-A receptor when compared to ac-CCK-7 (**2**) together with enhanced stability towards degradation.
Efforts to modify the ac-CCK-7 (**2**) compound at Abbott resulted in heptapeptide derivatives, A-71378 (**5**) and A-71623 (**6**) (Figure 5.3), with similar potency to Ro 23-2154 (**3**) and Ro 23-7014 (**4**), but with enhanced selectivity towards the CCK-A receptor (**5**: 1140 and **6**: 1216 versus **3**: 225 and **4**: 425). Astra-Zenaca published a series of peptide-type CCK-A agonists [166, 167].

[1] the numbers behind compound names are the compound numbers which are used in the figures and the tables.

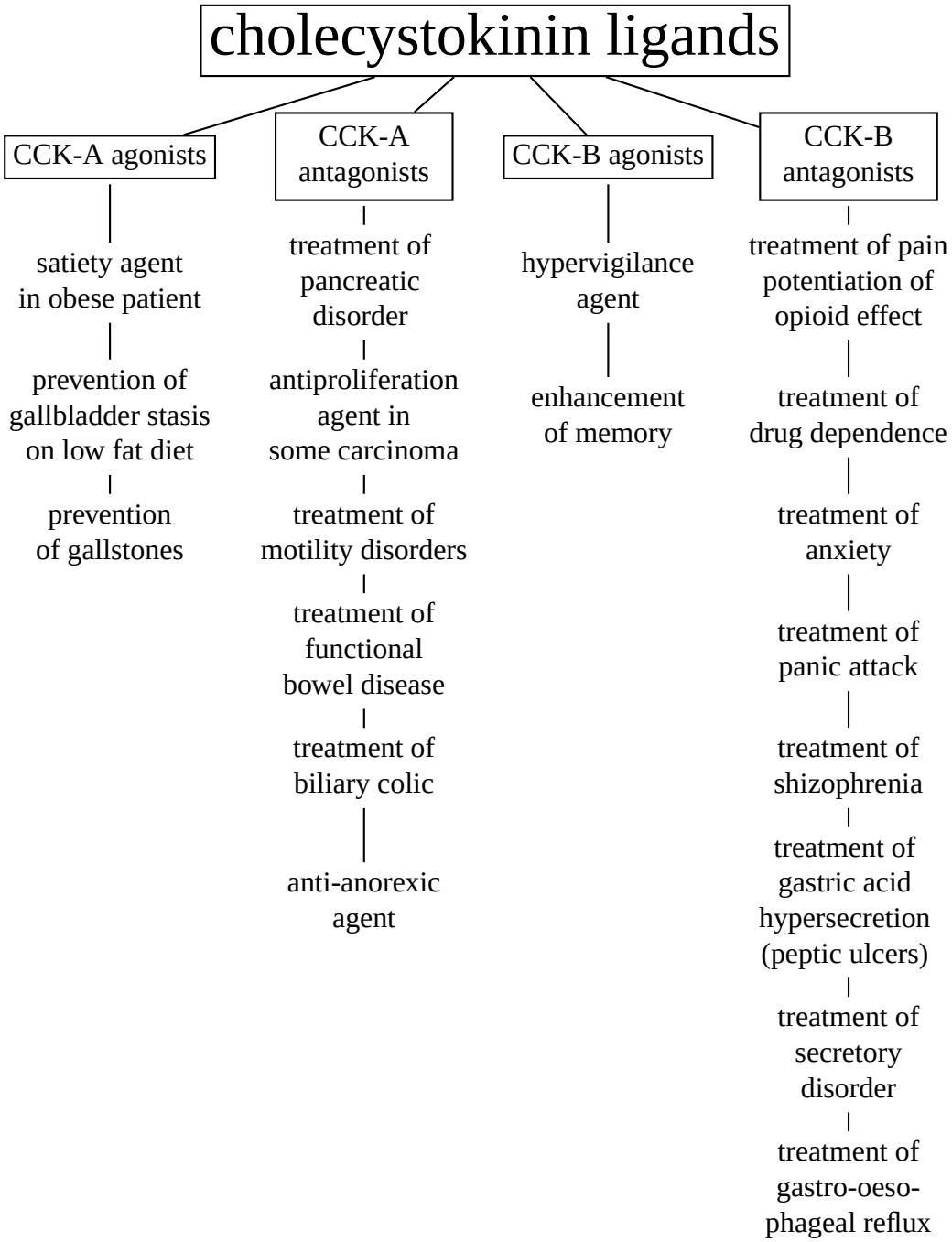


Figure 5.1: Potential therapeutic applications of CCK-receptor ligands [29].

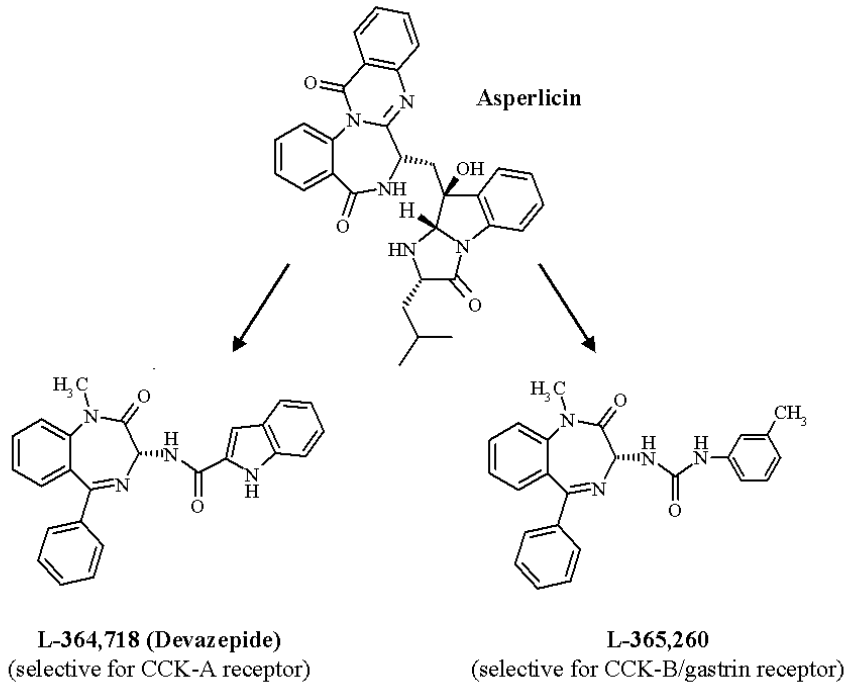


Figure 5.2: Example ligands for the CCK-A and CCK-B receptors.

Nr.	Name	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
1	(sulf.)CCK-8	0.28	0.64	[159]
2	Ac-(sulf.)CCK-7	0.60	4.4	[160]
3	Ro 23-2154	0.4	90	[161]
4	Ro 23-7014	0.4	170	[161]
5	A-71378 (Abbott)	0.5	570	[156]
6	A-71623 (Abbott)	3.7	4500	[156]
7	PD170292	1.2	6.7	[162]
8	SR146131 (Sanofi Recherche)	0.56	168	[163]
9	GW 7854	126	10	[164]
10	GW 5823	22.9	1000	[165]

Table 5.1: Selectivity profile of selected CCK-receptor agonists (Figure 5.3).

Only few studies report on the design of potent non-peptide agonists with substantially reduced peptide character as potential drugs for , e.g., the treatment of obesity. The compound, PD-170292 (**7**), was reported [162] as a simplified analogue of A-71623 (**6**) (both having a lysine bearing a phenylurea at the amino group). Remarkably PD-170292 (**7**) acts as an agonist towards the high-affinity and as an antagonist towards the low-affinity state of the CCK-A receptor, and as an antagonist towards the CCK-B receptor [162].

SR146131 (**8**) (Sanofi Recherché) [163, 168] can be envisioned as a peptidomimetic agonist which has even more structural similarities with the nonpeptide antagonist SR27897B (**11**) than with the endogenous peptide, it attempts to mimic. Obviously the attachment of a lipophilic

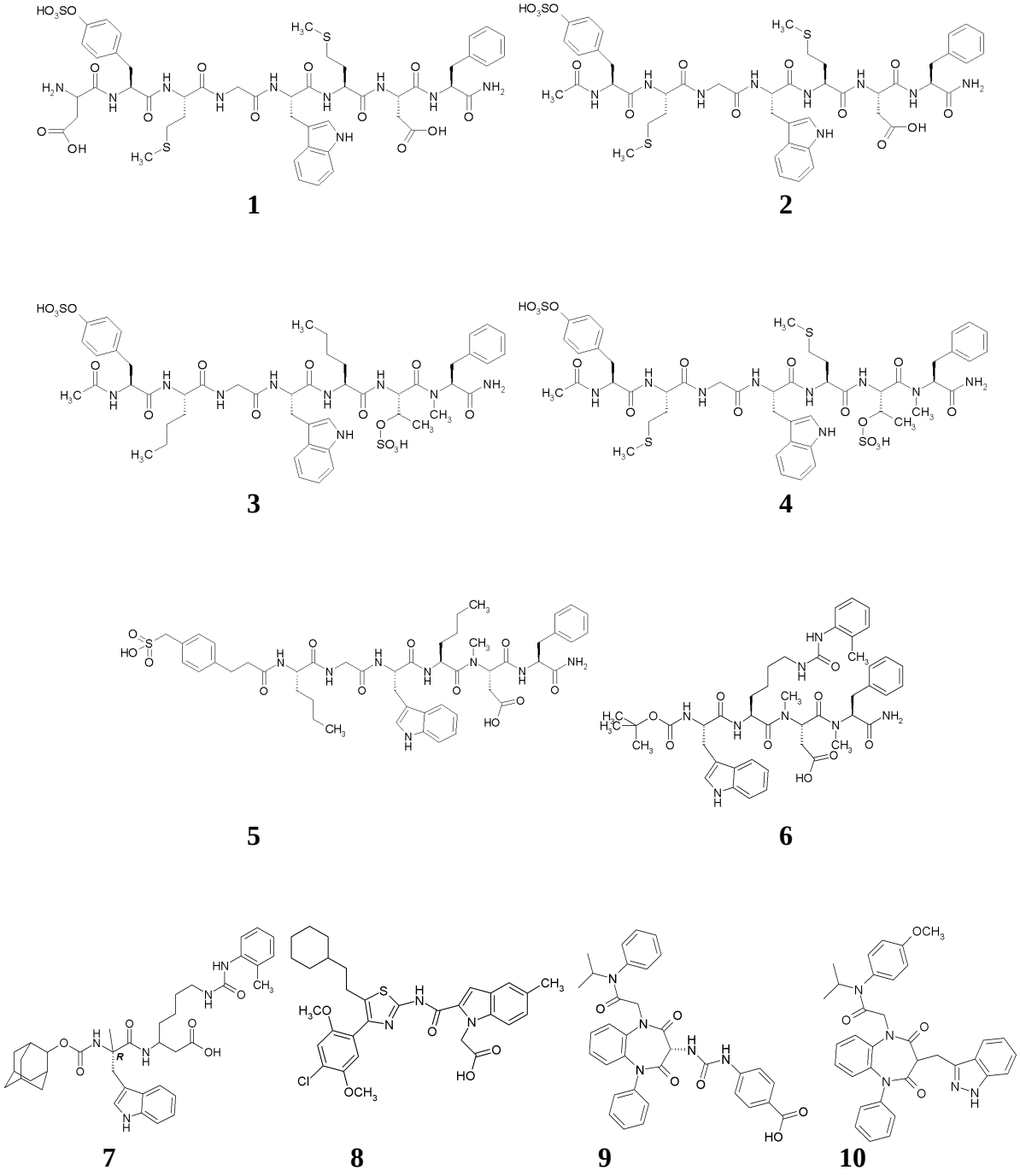
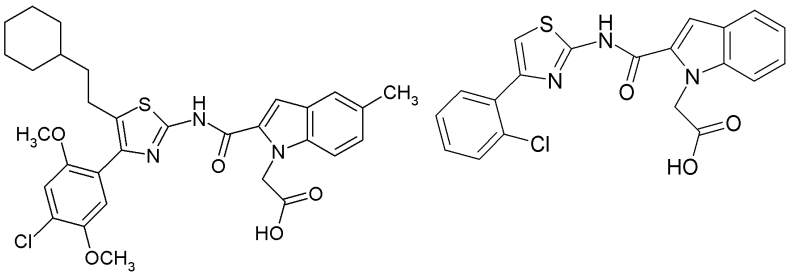


Figure 5.3: selected CCK-A agonists

cyclohexyl sidechain onto the thiazole ring interconverts an agonist (SR146131 (**8**)) to an antagonist (SR27897B (**11**)) (See Figure 5.4).



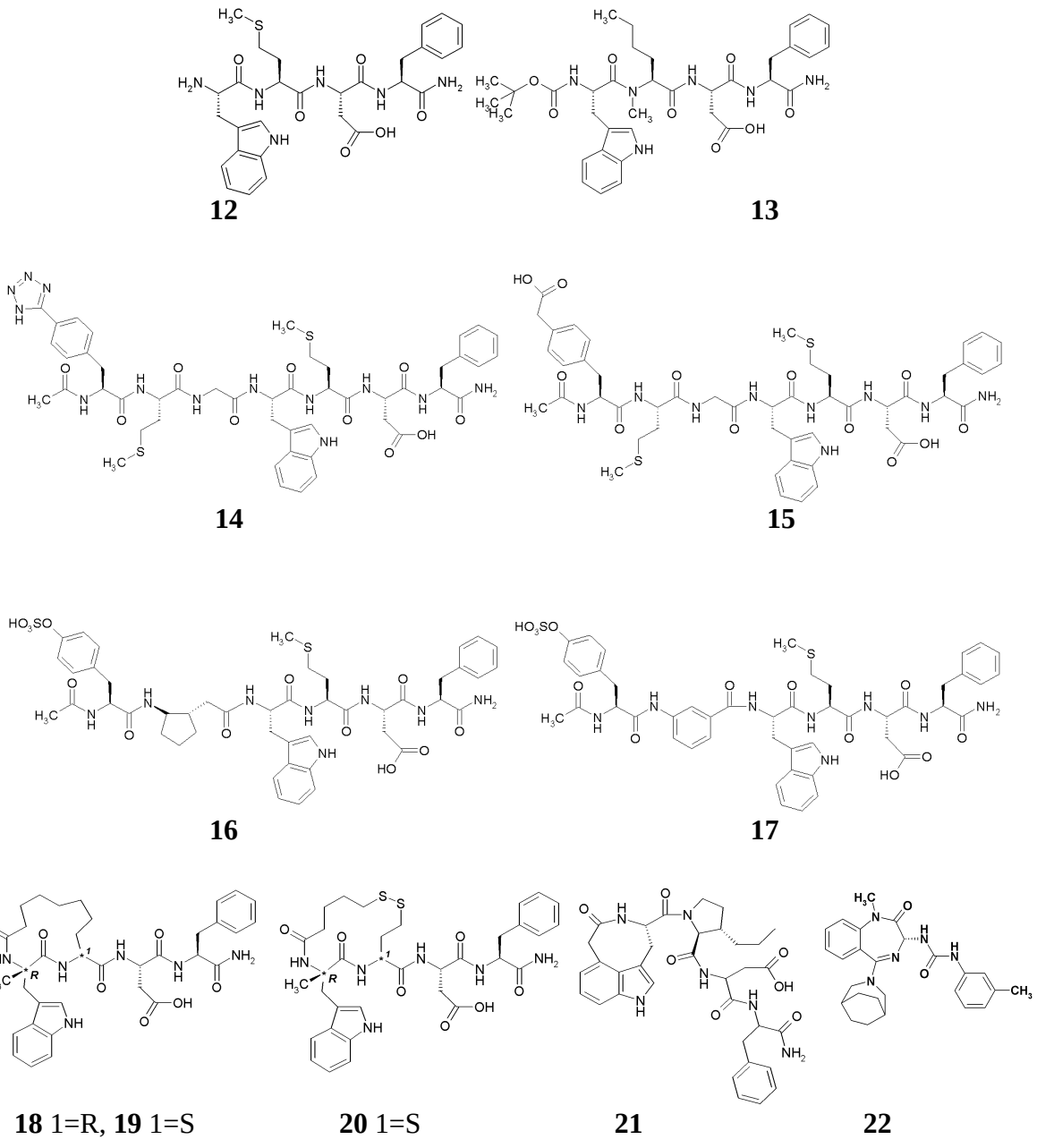
8 **11**

*Figure 5.4: Structural similarity between the agonist, SR146131 (**8**) and the antagonist SR27897B (**11**) [163, 168].*

Researchers at Glaxo Wellcome disclosed 1,5-benzodiazepines (e.g. GW 7854 (**9**) and GW 5823 (**10**)) with CCK-A receptor agonist activity by rationally altering the different side chains of 1,5-benzodiazepines [169, 170, 165, 171]. GW 7854 (**9**) has affinity for both CCK receptors, but it acts as CCK-A agonist and CCK-B antagonist. This mixed character is ideally suited for a potential drug candidate, since CCK-B agonist activity would be not desirable due to the potential peripheral (enhanced gastric acid secretion) or central (enhanced anxiety) side effects [164]. Also GW 5823 (**10**) shows CCK-B antagonist activity and even better CCK-A agonist activity. Despite this promising profile, pharmacokinetic studies carried out on rats revealed for this compound poor bioavailability and relative high clearance [165, 171].
This close structural similarity among nonpeptide agonists and antagonists suggests, the existence of similar binding pockets within the receptor protein, even though the steric overlap might be marginal. These two binding pockets are probably less distinct than the divergence between these two sites and the binding site of the endogenous peptide ligand. In summary, one can assume the existence of multiple binding sites within one receptor for the different classes of ligands. This hypothesis has to be considered when comparing antagonists with agonists in the context of docking studies of low-molecular weight ligands into the receptor protein.

5.3 CCK-B Agonists

Potent and selective CCK-B agonists were obtained by restricting the conformation of the C-terminus of the CCK peptide. Again a group at Hoffmann-La Roche rationally changed several positions of the CCK-7 peptide and derived various selective peptidic CCK-B agonists, **14**-**17**. (Figure 5.5) [172, 160].
Based on the finding that a CCK-4 peptide derivative, Boc-Trp-(NMe)Nle-Asp-Phe-NH_2 (**13**) [173], was found to be a potent and selective CCK-B agonist, cyclic CCK-B agonists were synthesised **(18**-**20** in Figure 5.5) [159]. Different bridge moieties between the N-terminus of the tetrapeptide derivative and the Nle/Met residue were designed. A disulfide bridge moiety in the cyclic peptide increased CCK-A receptor recognition. It is not surprising that the nature of the ring system has influence on the CCK-A versus CCK-B selectivity, because also the



12 **13**

14 **15**

16 **17**

18 1=R, **19** 1=S **20** 1=S **21** **22**

Figure 5.5: selected CCK-B agonists

Nr.	Name	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
1	(sulf.)CCK-8	0.28	0.64	[159]
12	CCK-4	> 35000	19.5	[159]
13	Boc-Trp-(NMe)Nle-Asp-Phe-NH_2	N.D.[1]	0.8	[173]
14	Tilley91-34	3.0	0.50	[172]
15	Tilley91-28	2.6	54	[172]
16	Tilley92-26	25	0.029	[160]
17	Tilley92-20	2.1	0.48	[160]
18	RB 360 I (1 = R)	N.D.[1]	4372	[159]
19	RB 360 II (1 = S)	2210	15	[159]
20	RB 380 (1 = S)	271	51	[159]
21	A-63387 (Abbott)	6300	0.7	[29]
22	L-740,093 (S)	N.T.[2]	19.5	[154]

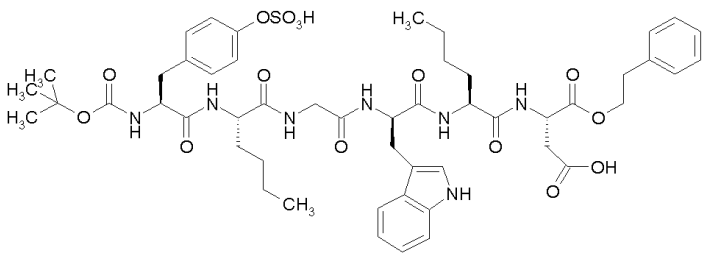
[1] N.D. = not determined, beyond the measurable boundary, [2] N.T. = not tested

Table 5.2: binding affinity for the two CCK-receptors of selected CCK-B Agonists (Figure 5.5).

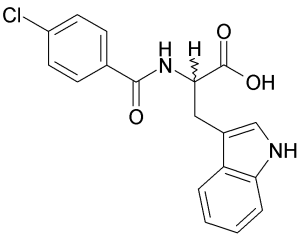
replacement of the Met residue in Boc-CCK4 by a Nle residue (Ro 23-2154 (**3**)**)** or by a lysine bearing a phenylurea in place of the amino group (A-71623 (**6**)), resulted in CCK-A agonists with enhanced selectivity. (Figure 5.3).
Abbott designed an even more restricted structure, A-63387 (**21**). This tetrapeptide (**21**) has a 9000 fold CCK-B selectivity over CCK-A and an affinity similar to that of CCK-8 (**1**). [29]
Due to their peptide or pseudopeptide nature, all these ligands show low bioavailabilities and unsatisfactory pharmacokinetics, which renders their application as orally-active drugs rather questionable.
A compound which might be not only a pharmacological tool, but also a therapeutic agent is L-740,093 (S) (**22**). L-740,093 (R) (**73**) is a potent CCK-B antagonist, however inversion of the stereogenic centre of the benzodiazepine moiety converts this ligand into a CCK-B agonist, L-740,093 (S) (**22**) [154] (Figure 5.5).

5.4 CCK-A Antagonists

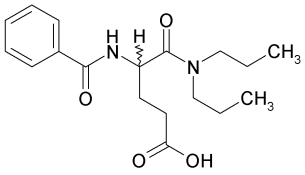
In many investigations a large number of CCK-A antagonists were developed. Modifications of the C-terminus of the peptide CCK were the basis for the search of peptide antagonists. Deletion of the C-terminal phenylalanine residue, inversion of L-Trp to D-Trp or changing the Met into a Nle residue, have resulted in, e.g., Boc-Tyr(SO_3H)-Nle-Gly-D-Trp-Nle-Asp-O-CH_2-CH_2-C_6H_5 (**23**) (Figure 5.6) [174]. However, this compound is only marginally selective for the CCK-A receptor.



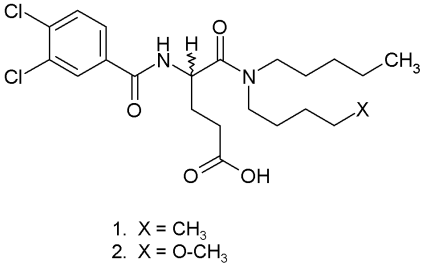
23



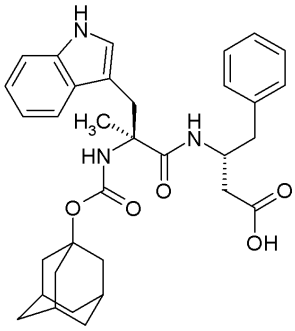
24



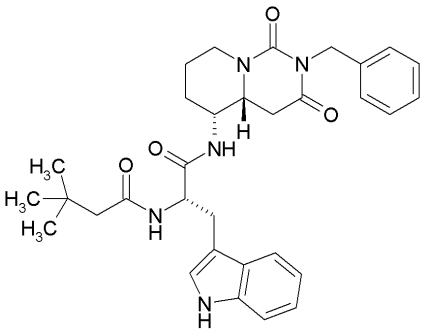
25



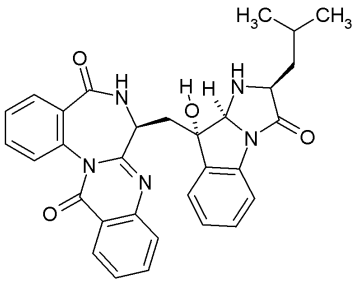
1: **26** 2: **27**



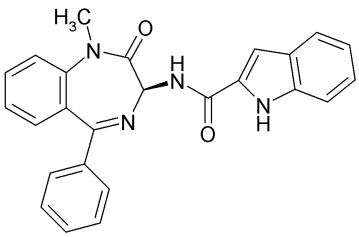
28



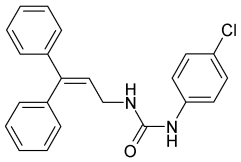
28



30



31



32

Figure 5.6: selected CCK-A antagonists

Nr.	Name	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
1	(sulf.)CCK-8	0.28	0.64	[159]
23	Boc-Tyr(SO_3H)-Nle-Gly-D Trp-Nle-Asp-O-CH_2-CH_2-C_6H_5	50	500	[174]
24	benzotript	84000	102000	[175, 115]
25	proglumide	6300	875000	[175, 115]
26	lorglumide (CR 1409)	130 (49)	300000	[176, 177] ([178])
27	loxiglumide (CR 1505)	330	9900	[176, 177]
28	CAM-1481 (PD 140,548)	2.82	260	[179]
29	IQM-95,333	0.62	>5000	[180]
30	asperlicin	1400	>100000	[115]
31	L-364,718 (Merck)	0.08	270	[115]
32	vdBent92-6	90	N.T.[1]	[178]

[1] N.T. = not tested.

Table 5.3: binding affinity of selected CCK-A antagonists for the two CCK-receptors (Figure 5.6).

Structural departure from the peptide character towards the development of peptidomimetic and nonpeptide antagonists revealed compounds with enhanced selectivity and acceptable pharmacokinetic profile. These CCK-A antagonists can be classified into four large families:

- peptidomimetic or peptoid family
- dipeptoid family
- tetrapeptide restricted family
- benzodiazepine family

The first family comprises well known members such as benzotript (**24**), proglumide (**25**), lorglumide (CR 1409) (**26**) and loxiglumide (CR 1505) (**27**) [181, 176, 177] (Figure 5.6), the last three compounds being potent and selective glutamic acid derivatives.
Extensive SAR studies were derived on proglumide and its analogues. It turned out that the R-isomers were more potent than the corresponding enantiomers. This influence on the stereochemistry parallels the findings of stereoisomerism of the Trp position. Loxiglumide (**27**) and Lorglumide (**26**) display increased affinity and selectivity towards the CCK-A receptor. These compounds have been evaluated by Rotta Labs in clinical studies [29]. However, only proglumide (**25**) (Milide $^{(tm)}$ or Milid $^{(tm)}$) has been marketed and is used for treating gastric ulcers.
The structures of the dipeptoid family were based on the two most important amino acid residues of CCK-8 and CCK-4, namely tryptophan and phenylalanine. This research led to potent and selective CCK-A receptor antagonists such as CAM-1481 (PD 140,548, **28**, see Figure 5.6) [179].
As mentioned before, rigidification of the tetrapeptide structure in order to mimic the bioactive conformation of the C-terminus of CCK-8 turned out to be an important direction in the search of novel CCK-A antagonists, exemplified with the 1,3-dioxoperhydropyrido[1,2-c]pyrimidine derivatives, IQM-95,333 (**29**) being the most potent member of this series [180].

One of the most important findings in the development of CCK-A and CCK-B ligands was the discovery of the CCK-A antagonist activity of asperlicin (**30**) by receptor-based screening technology at Merck (Figure 5.6). Asperlicin (**30**) is a fermentation product from *Aspergillus alliaceus* [115]. This discovery paived the route towards the design of nonpeptide antagonists and agonist for both CCK-receptors. Efforts to simplify its structure, to enhance affinity and oral activity, stimulated Merck scientists to design, synthesis and evaluate several series of compounds with a hybrid structure encompassing the 1,4-benzodiazepine part and an aromatic part (for example ref. [182, 183, 184, 185, 186, 187]). These hybrid structures were designed to mimic the asperlicine structure [188]. L-364,718 (devazepide) (**31**) was the best compound of these benzodiazepine derivatives for the CCK-A receptor with a high (0.08 nM) binding affinity and high CCK-A receptor selectivity (3375 fold) . Devazepide (**31**) has been used in the last few years for enhancing our understanding of CCK receptor location, function and pharmacological actions. The benzodiazepine family emerged over the last decade to one of the largest and most important families of the CCK-A antagonist types.
On the basis of the structural homology between lorglumide (**26**) and L-364,718 (**31**), a novel series of nonpeptide CCK-A antagonists was designed and synthesised [178]. Compounds of this series (Figure 5.6) are clearly nonpeptide antagonists, since the peptidic character is completely abolished.

5.5 CCK-B Antagonists

CCK-B antagonists cluster in different compound classes, the most important members will be introduced in the following sections.

5.5.1 Glutamic Acid Structure

Modifications of CCK-A antagonist Lorglumide (**26**) (Figure 5.6) led to the specific CCK-B antagonist spiroglumide (CR2194) (**33**), with only a 11 fold selectivity enhancement for the CCK-B receptor [189]. Further modifications led to CR2622 (**34**) with higher affinity (70 fold) and better CCK-selectivity (369 over 9.6) (Figure 5.7) [190]. The authors rationalise this affinity of CR2622 (**34**) by its capacity to mimic the bioactive conformation of the terminal gastrin pentapeptide. The low-energy conformation of CR2622 (**34**) superimposed well on the low-energy pentagastrin structure (see Figure 5.8 for a possible superpositioning).

5.5.2 Tryptophane Dipeptoid Derivatives

Comparable to the CCK-A antagonist programs, a series of tryptophane and phenylalanine derivatives were synthesised and evaluated as putative CCK-B antagonists by Parke-Davis. One of the most potent compounds is CI-988 (PD 134,308) (**35**) [191], which suffers from a poor bioavailability. This clinical finding was attributed to poor absorption and efficient hepatic excretion. Consequently, compounds were developed with reduced molecular weight, that should lead to better absorption. In this process the key moieties, α-methyltryptophan and adamantyloxycabonyl, were kept intact, while the C-terminus was extensively modified. CI-1015 (**36**) emerged as a good CCK-B antagonist with excellent CCK-B selectivity. Its bioavailability in rat

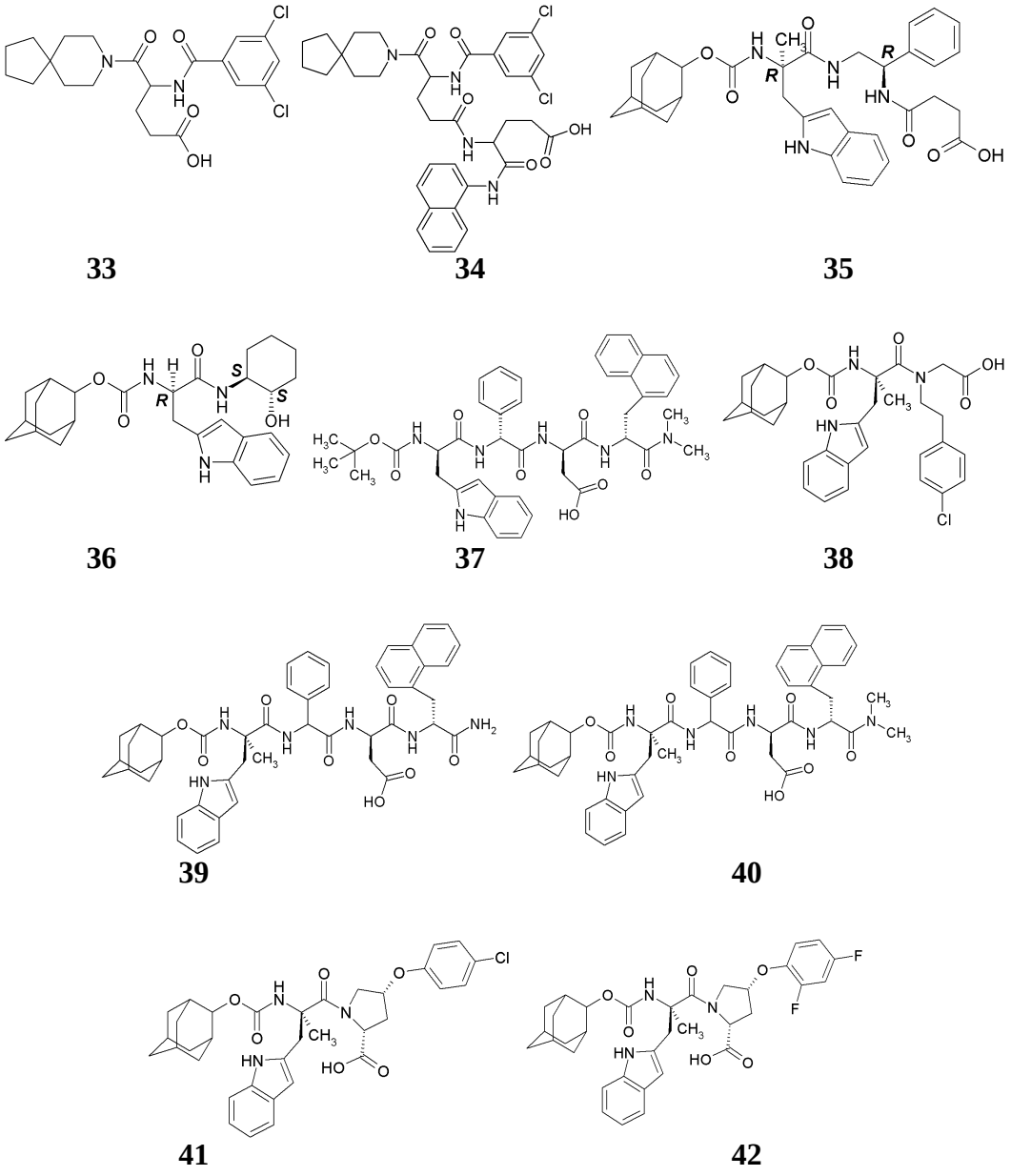


Figure 5.7: selected glutamic acid and tryptophane dipeptoid CCK-B antagonists

is improved by 10 fold over CI-988 (**35**) and also the blood-brain permeability was enhanced relative to CI-988 (**35**), which rendered CI-1015 (**36**) as a promising development candidate. [192]

The French research group of B.P. Roques designed potent and selective dipeptoid CCK-B antagonists by incorporating non-natural hydrophobic amino acids together with conformational restrictions achieved by proline [195, 196]. Affinity and selectivity of these analogues was probed systematically. As similarly observed for the CCK-A antagonists chirality of the Trp residue plays in all six structures (**37**-**42**) an important role in the binding affinity [195, 196, 194, 193]. In contrast, the two other stereogenic centres of compounds **39** and **40** possess significantly less influence on the binding affinity [195]. Remarkable is the fact that the bind-

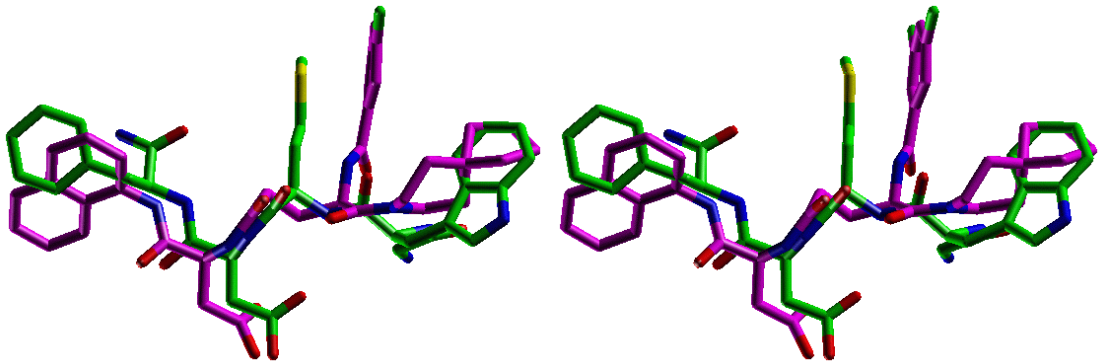


*Figure 5.8: Stereoview of a possible superpositioning of the CCK-B antagonist CR22 (**34**, magenta) and the terminal gastrin pentapeptide (green).*

Nr.	Name	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
1	(sulf.)CCK-8	0.28	0.64	[159]
33	CR 2194 (spiroglumide)	13500	1400	[189]
34	CR 2622	7380	20	[190]
35	CI-988 (PD 134,308)	4300	1.7	[192]
36	CI-1015	2900	3.0	[192]
37	Boc-Trp-Phg-Asp-1Nal-N(CH_3)$_2$	1000	39	[193]
38	^{2}Adoc-D-αMeTrp-N-[2-(4-chlorophenyl)ethyl]glycine (RB211)	1060	6.1	[194]
39	^{2}Adoc-D-αMeTrp-Phg-Asp-NalN(CH_3)$_2$ (Weng96-13)	54	3.5	[195]
40	^{2}Adoc-D-αMeTrp-Phg-Asp-NalNH_2 (Weng96-16)	250	3.4	[195]
41	^{2}Adoc-D-αMeTrp-D-cis-Hyp(Ph-pCl)-OH (Bellier97-6e)	1560	24	[196]
42	^{2}Adoc-D-αMeTrp-D-cis-Hyp(Ph-o,pF_2)-OH (Bellier97-11e)	394	17.6	[196]

Table 5.4: binding affinity of selected glutamic acid and tryptophane dipeptoid CCK-B antagonists for the two CCK-receptors (Figure 5.7).

ing affinity of the restricted compound **41** is a factor 2 lower than RB211 (**38**), but it is 3 times more potent in the inositol phosphate formation induced by CCK-8 (a common CCK-antagonist activity experiment) [196]. Compound **42** on the contrary has similar binding affinity as compound **41**, but has a similar potency as RB211 (**38**) [196]. The only difference between these two structures are the substituents on their phenyl ring. On the basis of these results, it can be suggested that there exist high-affinity and high-potency antagonists (**41**) and high-affinity and low-potency antagonists (**38** and **42**) [196], whereby the structures do not differ significantly. This could imply that there are two CCK-B receptor states of distinct affinity as is suggested by Bellier et al. [196]. However, slightly different CCK-B antagonist binding sites or different binding modes of the ligands can also explain these results.

It is also remarkable that, although the bindings affinities towards the CCK-B receptor of compounds **39** and **40** are very similar, their capacity to influence the inositol phosphate formation is different. Compound **39** functions as a full CCK-B antagonist with similar potency as CI-988 (**35**) and L-365,260 (the prototype CCK-B antagonist), while compound **40** is a partial agonist. Furthermore, compound **40** is a partial agonist in the IP formation, an agonist by pentagastrin acid secretion and an antagonist in the firing rate of rat CA2 hippocampal neurons (CCK-B

agonist have shown to stimulate the firing rate of these neurons in a dose-dependent manner). The only difference in the structure of these two compounds is the more bulky C-terminus of compound **40** [195].

5.5.3 Quinazolinone Series

A research group at Lilly [197] (successfully) focused on alternative substructures embedded within the molecular framework of the natural product asperlicin (**30**). These substructures provided a rational starting point for the effort to design structurally novel series of nonpeptide CCK receptor ligands derived from asperlicin (**30**). By disconnecting several bonds of the asperlicin structure, two compound classes were obtained with reduced chemical complexity (Figure 5.9). The first class refers to the tryptophan-derived CCK-ligands, as for example the CCK-A antagonist benzotript (**24**). The second class describes novel compounds, notably quinazolinones. An advantage of this structural class is that it lacks the asymmetric centres found in, e.g., the tryptophan dipeptoid derivatives (**35**-**42**) and asperlicin (**30**). By synthesising and evaluating a series of quinazolinones [197, 198, 199], the influence of the substitution pattern of this heterocyclic system was investigated. It is remarkable that all these structures display excellent CCK-B receptor selectivity, whereas the natural product asperlicin (**30**) is a selective CCK-A receptor antagonist.
Parke-Davis [200, 201] developed a new series of quinazolinone derivatives by combining key fragments of the Lilly and Merck series (Figure 5.9). They investigated the indole moiety (replaced by a phenyl ring with different substituents), introduced an urea functionality in the linker between the quinazolinone ring and the indole moiety and investigated the optimal substitution pattern at the phenylring attached to the quinazolinone N-3.
A small fraction of compounds synthesised by Lilly and Parke-Davis is given in Table 5.5. Special emphasis is put on the elaborated structure activity relationship, since various low-molecular weight antagonists used for our docking study (Chapter 6) were selected from this series of congeneric compounds.
Clearly, the 3-isopropoxyphenyl is the preferred substitution on the N-3 quinazolinone. Furthermore, the CCK-B affinity is enhanced by electron-withdrawing substitutions on the terminal phenyl ring attached to the urea moiety. The perfect linker is -NH-NH-CO-NH [200, 201]. This optimised linker has good structural similarity with the structure of asperlicin (**30**) (Figure 5.10), although it is a longer linker as previously anticipated by the structural comparison in Figure 5.9.

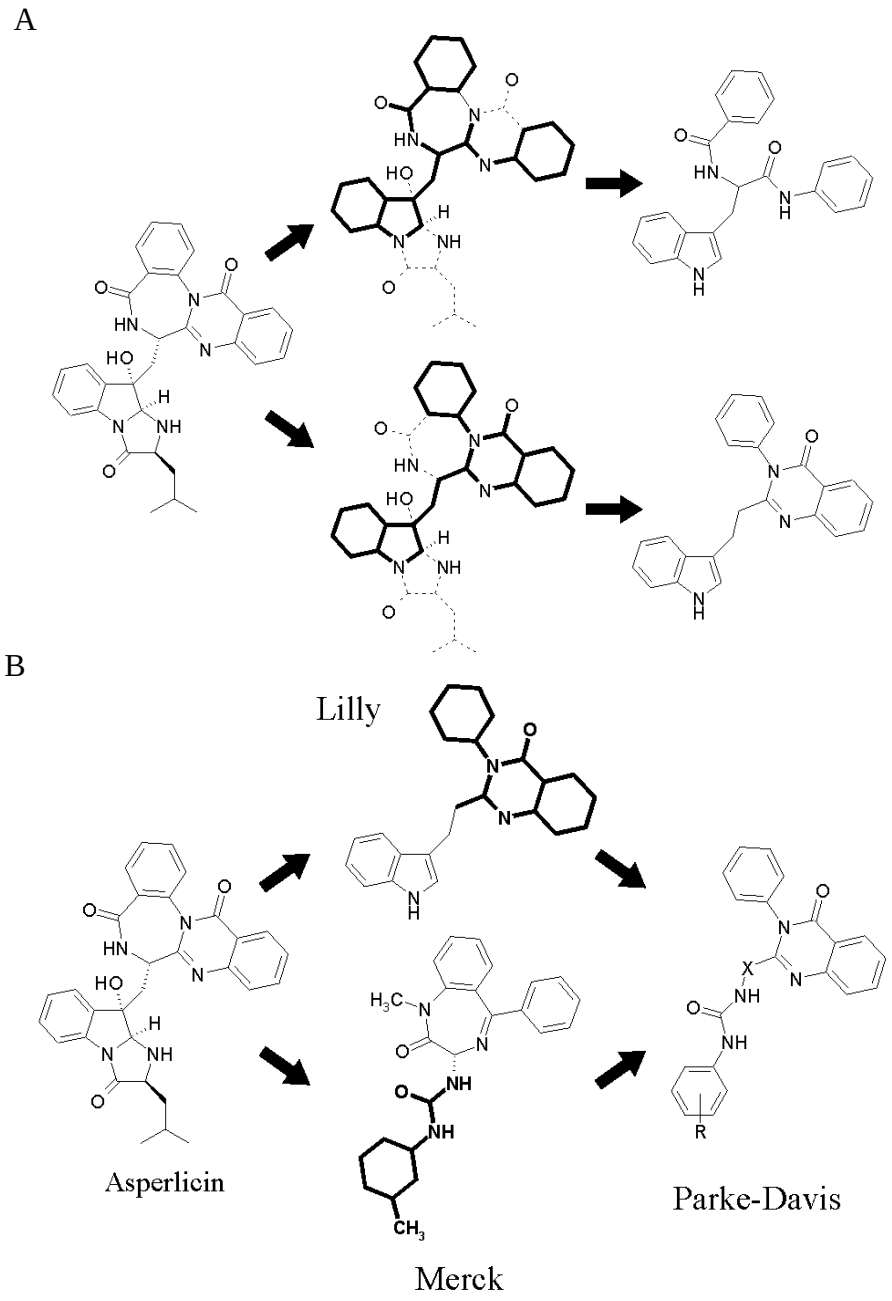
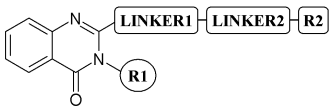
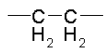
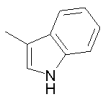
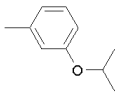
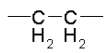
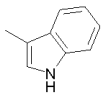
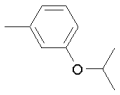
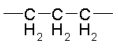
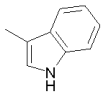
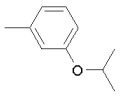
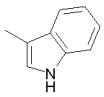
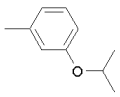
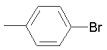
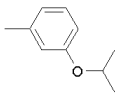
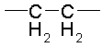
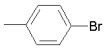
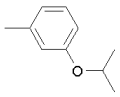
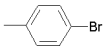
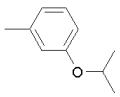
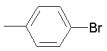
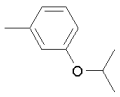
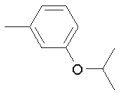
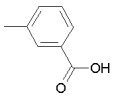
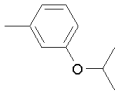
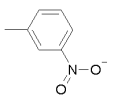


Figure 5.9: (A) Two possible bond disconnection strategies of the asperlicin structure. The three aromatic domains are preserved and the chemical complexity is reduced. The first (above) yields a substructure common to tryptophan derivatives of CCK-receptor ligands. The second (below) yields the novel series of 3-phenyl-4(3H)-quinazolinone structures. (B) Further development by Parke-Davis by combining key fragments of the Lilly and the Merck series.

Table 5.5: Binding affinities of selected quinazolinone derivatives for the CCK-receptors.



Nr.	R1	Linker 1	Linker 2	R2	CCK-A [nM]	CCK-B [nM]	ref.
43		-			N.D. [1]	670	[197]
44 LY-247348		-			N.D. [1]	26	[197]
45		-			N.D.[1]	>1000	[198]
46		-			N.D.[1]	>1000	[198]
47		-			>1000	>1000	[200]
48					1630	585	[200]
49					984	137	[200]
50					3430	14	[200] [201]
51					17300	108	[200]
52					8110	29	[201]
53					2560	1.0	[201]
54					6400	11	[201]
55					N.D. [1]	97	[201]
56					4900	63	[201]

[1] N.D. = not determined, beyond measurable boundary.

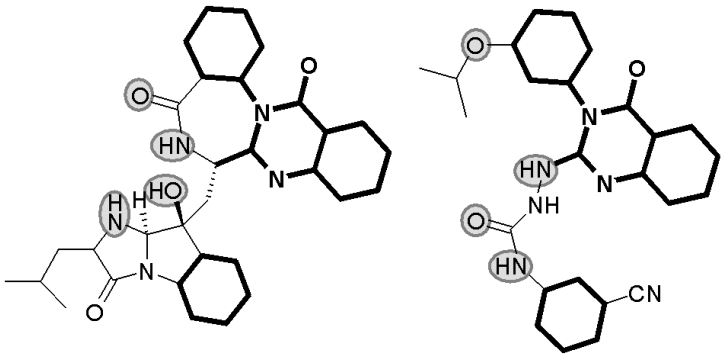


Figure 5.10: Similarity between asperlicin and quinazolinone derivative with optimised urea linker.

5.5.4 Diphenylpyrazolidinone Series

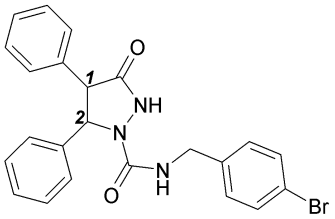
Following a random screening approach, Lilly identified a pyrazolidinone derivative as CCK-B receptor ligand. Again, the stereochemistry of the compounds **57**-**62** plays a crucial role in the binding affinity of the compounds (Figure 5.11).

5.5.5 Ureido-Acetamide Derivatives

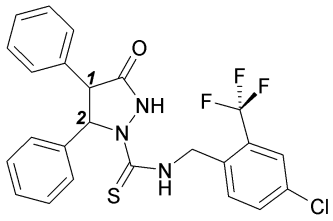
Bertrand et al. [202] at Rhône-Poulenc Rorer developed a further family of non-peptide CCK-B receptor antagonists, notably the ureido-acetamide family. The three representative members of this family, RP 69758 (**63**), RP 71483 (**64**) and RP 72540 (**65**) (Figure 5.11), displayed nanomolar affinity for the CCK-B receptor with decreased binding to the CCK-A receptor. *In vitro* tests demonstrated these three compounds (**63**-**65**) to be potent CCK-B antagonists *in vivo*. The ureido-acetamides do not interact with a wide array of other neurotransmitter or hormone receptors, such as histamine H_1, dopamine D_2, neuropeptide Y, neurotensin, serotonine 5-HT_2, somatostatin, tachykinin NK_1, and bradykinin receptors, rendering them as highly selective drug candidates [202].

Nr.	Name	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
1	(sulf.)CCK-8	0.28	0.64	[159]
57	LY 288512 (Lilly)	6400	370	[156]
58	LY 288513 (Lilly)	20500	19	[156]
59	LY 262691 (Lilly)	11600	31	[156]
60	LY 294920 (Lilly)	17	1900	[156]
61	LY 294919 (Lilly)	810	550	[156]
62	LY 219057 (Lilly)	42	880	[156]
63	RP 69758 (Rhône-Poulenc Rorer)	1254	9.0	[202]
64	RP 71483 (Rhône-Poulenc Rorer)	2338	2.4	[202]
65	RP 72540 (Rhône-Poulenc Rorer)	164	0.76	[202]

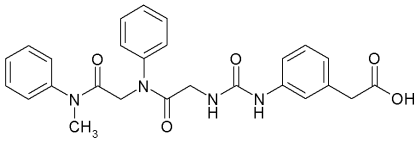
Table 5.6: Bindings affinities of selected diphenylpyrazolidone and ureido-acetamide CCK-B Antagonists for the CCK-receptors (Figure 5.11).



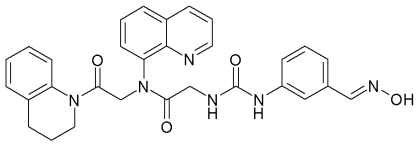
57=RS, **58**=SR,
59=racemic trans



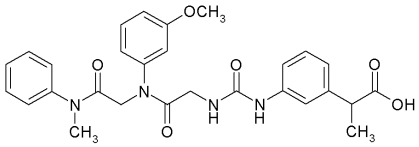
60=RS, **61**=SR
62=racemic trans



63



64



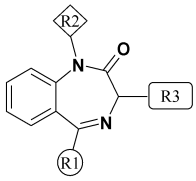
65

Figure 5.11: selected diphenylpyrazolidinone and ureido-acetamide CCK-B Antagonists.

5.5.6 Benzodiazepine Family

As described in Section 5.4, the selective and potent CCK-A antagonist, L-364,718 (**31**), was designed from the natural product asperlicin (**30**) [115]. In the development of L-364,718 (**31**) hundreds of analogues of this benzodiazepin derivative were synthesised and evaluated. These compounds revealed the information that subtle structural modifications resulted in loss of receptor subtype selectivity. Replacing the indole substituent by a 4-chlorobenzoyl group (**67**) revealed an equivalently potent and selective derivative as L-364,718 (**31**). Subsequent exchange of the N-1 methyl substituent by an ethoxycarbonyl group, or the 3-amide linkage by an urea and the N-1 methyl substituent with an H-atom decreased the CCK-A affinity substantially while the CCK-B affinity increased slightly. Combination of these two features (N-1 ethoxycarbonyl substituent and urea linkage) resulted in the first nonpeptide derivative (**69**) with a selective nanomolar potency for the CCK-B receptor (Table 5.7). Two other compounds with a urea linkage, the N-1 methyl-4-chlorobenzoyl derivative and the N-1 methyl-3-methylbenzoyl derivative (**70**), have a slightly decreased potency (23 and 7.1 nM versus 1.0 nM of the N-1 hydrogen-4-chlorobenzoyl derivative (**69**)). Again, stereochemistry is the reason for target discrimination since in both instances, the enantiomers with the 3S-configuration (**71**) showed selectivity for the CCK-A receptor whereas the 3R-configuration (**72**) showed selectivity for the CCK-B receptor. This is in contrast to the enantiomers of the N-1 methyl-4-chlorobenzoyl derivative (**67** and **68**) with the amide linkage, which are both highly selective for the CCK-A receptor. The 3R-enantiomer of the N-1 methyl-3-methylbenzoyl derivative is the well known L-365,260 (**72**), the prototype nonpeptide antagonist for the CCK-B receptor.

Table 5.7: Binding affinities of selected benzodiazepine derivatives for the CCK-receptors.



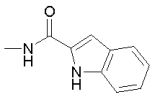
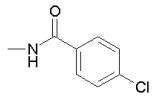
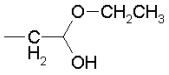
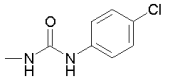
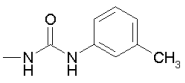
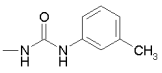
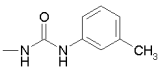
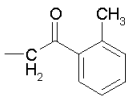
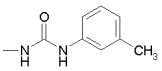
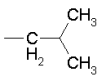
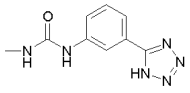
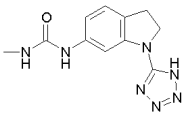
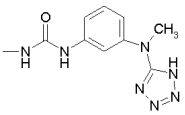
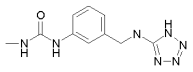
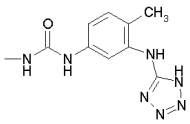
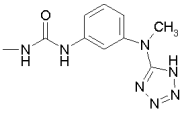
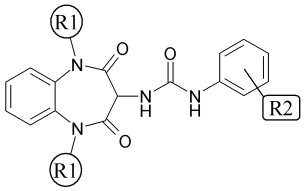
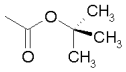
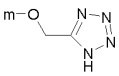
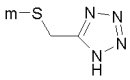
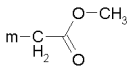
Nr.	R1	R2	R3	Stereo	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
31 L-364,718				S	0.08	270	[184]
66				R	8.3	3700	[203]
67				S	0.39	2900	[184]
68				R	2.9	11000	
69				RS	370	1.0	[197]
70				RS	8.1	7.1	[184]
71				S	3.0	151	
72 L-365,260				R	280	2.0	
73 L-740,093				R	1604	0.1	[204]
74 L-708,474				R	1820	0.28	[204]
75 YM-022				R	150	0.11	[158] [155] [29]
76 L-368,935				R	1434	0.14	[158] [155] [29]
77 L-738,425				R	4080	0.11	[205]
78				R	3000	0.58	[205]
79				R	3000	7.5	[205]
80				R	3000	1.1	[205]
81 L-736,380				R	400	0.054	[205]
82				R	480	137	[205]
83				R	3090	1.5	[205]

Table 5.8: Binding affinities of selected 2,4-dioxo-1,5-benzodiazepine derivatives with a plane of symmetry [206] for the CCK-receptors.



Nr.	R1	R2	CCK-A affinity [nM]	CCK-B affinity [nM]	ref.
84		m−CH_3	860	54	[206]
85		m−CH_3	380	18	[206]
86		m−CH_3	>10000	>1000	[206]
87		m−CH_3	2400	480	[206]
88		m−CH_3	860	23	[206]
89		m−CH_3	23	8	[206]
90		m−CH_3	100	2	[206]
91		m−CH_3	11000	>1000	[206]
92			>10000	84	[206]
93			2900	8	[206]
94			4400	16	[206]
95			1800	4	[206]
96		m−CH_3	500	2	[206]
97			1300	1	[206]
98			2500	48	[206]
99			>10000	>1000	[206]
100			2600	26	[206]

Due to the low bioavailability the development of a second generation of benzodiazepine CCK-B antagonists was pursued by scientists at Merck and Yamonouchi. These new compounds (**73-83**) were designed to show high potency and selectivity, which by adding of a water solubilising group should improve bioavailability as exemplified by L-740,093 (**73**). A selection of these compounds are given in Table 5.7. Again it is attempted to give an impression of the SAR of these derivatives, since the most selected compounds were used for our docking studies (Chapter 6).
It should be emphasised that the 2,4-dioxo-1,5-benzodiazepines (**84-100**) (Table 5.8) possess a plane of symmetry and therefore do not possess a stereogenic centre at the C3 position [206].

5.6 Receptor Binding for CCK-A and CCK-B

Since the benzodiazepine were first developed as CCK-A antagonists, extensive QSAR analyses and pharmacophore modelling have been applied to this compound class.
Based on molecular shape and QSAR analysis, Tokarski et al. [207] derived an 3D-pharmacophore model for the 3-amino substituent of the benzodiazepine CCK-A antagonist series. This pharmacophore model shows that the lipophilicity of the substituent in the para position increases the binding affinity, which is in contrast to the ortho substituent on the 3-amino phenyl ring [207]. However, there is a trade-off between size and lipophilicity. Compounds with atoms extending above and/or below the plane of the aromatic ring of the 3-amino substituent have a significantly lower activity. Therefore, Tokarski [207] suggested that there must be a non-allowed volume

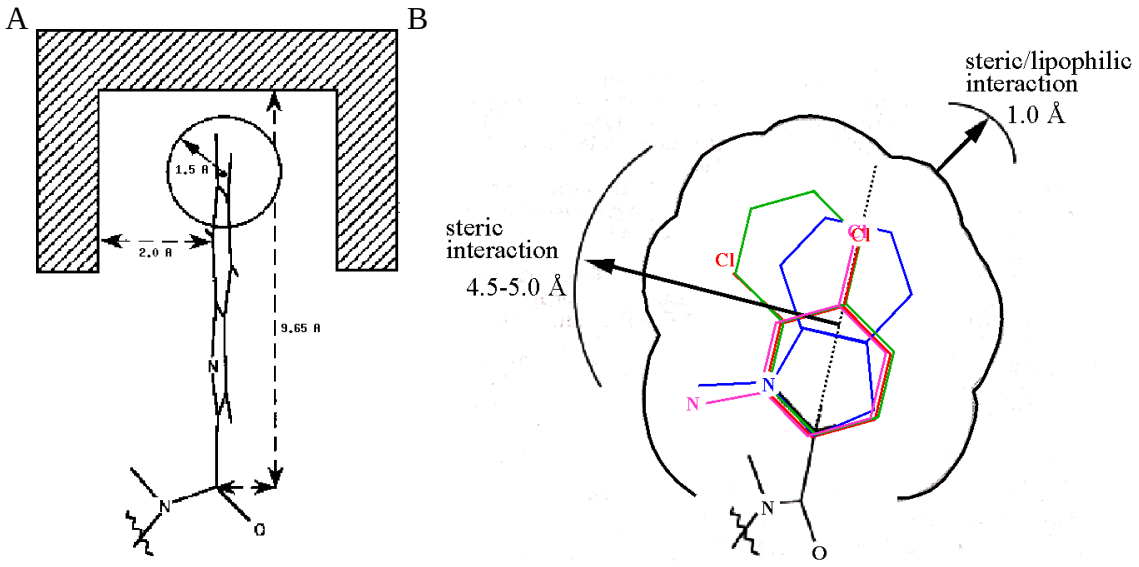


Figure 5.12: 3D-pharmacophore model for the 3-amino substituents of the benzodiazepine CCK-A antagonists developed by Tokarski et al. [207] (A) 1.3-2.0 Å above and below the aromatic ring-plane and 9.65 Å from the carbonyl carbon is the receptor limiting the accessible volume of the substituents. A hydrophobic sphere of 1.5 Å at the end of the pocket could be significant for a good binding affinity. (B) Planar view of the accessible volume depicted with three example substituents (3,4-di-Cl-phenyl - red; 2-naphthyl - green; o-NH_2-p-Cl-phenyl - magenta and 2-indolyl - blue).

about 2.0 Å above and below the plane of the aromatic ring, causing steric interaction between the receptor and the antagonists. Additionally, the receptor-excluded volume is defined at 9.65 Å from the carbonyl carbon, thus limiting the size of the substituents. The steric bulk of the 3-amino substituent is also restricted to 4.5 - 5.0 Å from the axis through the para substituent (see Figure 5.12). Accordingly, only a limited volume is available for the 3-amino substituent, giving rise to a constrained conformation of the aromatic ring. Therefore, the angle of the aromatic ring towards the plane of the 3-amino moiety is of crucial importance [207]. This angle of the aromatic ring (phenyl or indole) is depending on the physico-chemical properties of its substituents [208].

Furthermore, the substituents in the para position should be hydrophobic, since the para-OH and the para-Cl have only poor and moderate activity. Therefore, a hydrophobic sphere of 1.5 Å around the Iodinium-atom at the para position of the phenyl ring has to be considered in the 3D pharmacophore.

The fact that the 3-amino substituents of the benzodiazepine CCK-A antagonist have a limited bulk tolerance was also suggested by Sinha et al. [209] and van der Bent et al. [208]. Sinha postulated that the NHCO is favourable for the CCK-A antagonist, but the NHCONH-moiety is mediating CCK-B selectivity. The additional NH should perform a sterically unfavourable interaction with the CCK-A receptor, whereas it introduces an additional hydrogen bonding facility in the CCK-B receptor. Furthermore, they concluded that CCK-B binding of the benzodiazepine compounds is significantly depending on logP, suggesting that the compounds should bind in a large lipophilic binding pocket. In contrast to the CCK-A receptor, this pocket should be more flexible and sterically less constrained [209].

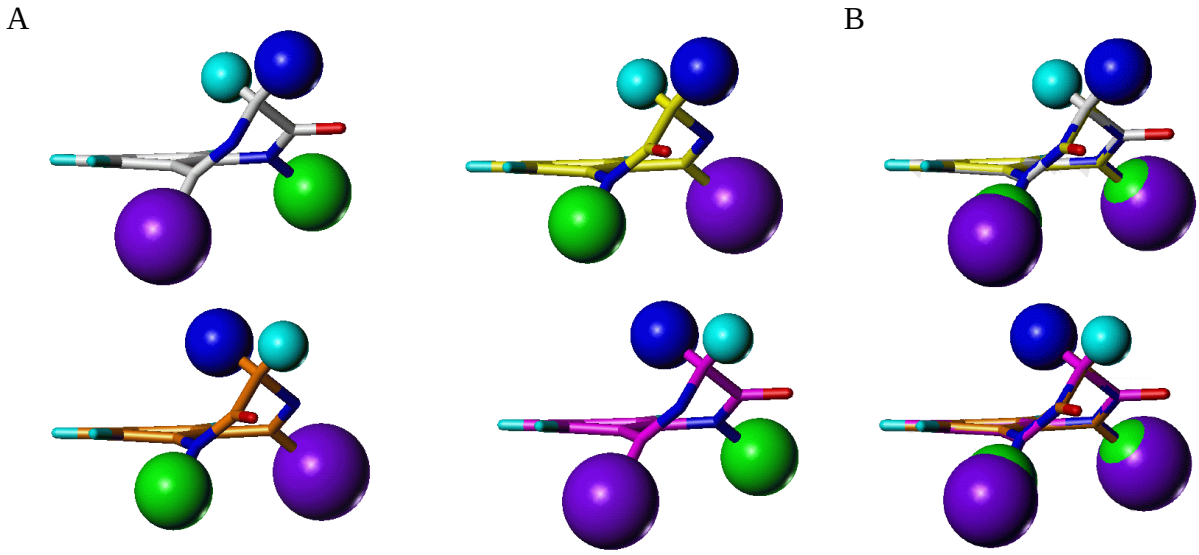


Figure 5.13: Benzodiazepine-derivatives in different conformations; (A) white structure: R isomer with the dark blue side chain in the pseudo-euatorial orientation; orange structure: R isomer with pseudo-axial; yellow structure: S isomer with pseudo-equatorial; and magenta structure: S isomer with pseudo-axial. (B) the dark blue side chains of the R-equatorial (white) and the S-equatorial (yellow) overlap however their purple and green side chains are pointing in opposite directions; the same is true for the R-axial (orange) and the S-axial (magenta).

Van der Bent et al. [208] suggested that the relatively low affinity difference between R and S isomers of the benzodiazepine CCK-A antagonists could be rationalised by the mutual sterical overlap of the two enantiomers in which both adopt either pseudo-axial or -equatorial side chain orientations (Figure 5.13). In this superposition, the bulky C5-phenyl rings of the two enantiomers are pointing in opposite directions (see Figure 5.13). From this conclusion it can further be postulated that the pockets on both sides of the benzodiazepine rings are relatively large and could even exhibit a certain symmetry, thus explaining the binding affinity of the symmetrical 2,4-dioxo-1,5-benzodiazepines. Although the slight preference for the S-isomer suggests that the pockets could have minor differences. With this data it can not be determined whether the benzodiazepine CCK-A antagonist have either pseudo-axial or -equatorial orientation. Experimentally determined structures of benzodiazepine compounds generally adopt the pseudo-equatorial orientation (Figure 5.14), but calculations of van der Bent [208] demonstrated a preference for the pseudo-axial orientations (orange and green structures of Figure 5.13). However, these calculations were performed in vacuum neglecting any explicit solvent

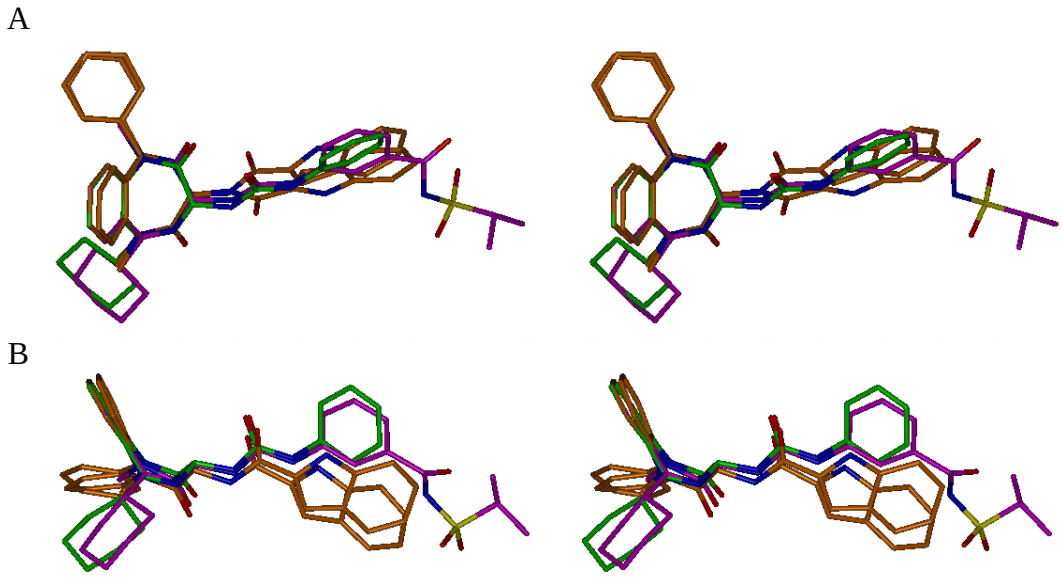


Figure 5.14: Stereo top and sideview of the superposition of the structures found in the CSD-database with a benzodiazepine ring and an NHCO-moiety; the entries SABKJEK (2x, orange), PIHFOB (green) and ZOVJUP (magenta).

environment, thus the resulting conformations might be artefacts, since vacuum-derived conformations tend to adopt an overall compact, folded conformation. In our docking study described in Chapter 6 it has been investigated whether the pseudo-axial or the pseudo-equatorial orientations are preferred in the binding site.

Chapter 6

Antagonist Binding Site

Identification of Antagonist Binding site: A Molecular Docking Approach.

"The major discriminating force between different molecules that bind to a single receptor is repulsion. This is another way of saying that the most important feature is fitting in the simple geometrical sense."

- Graham Richards

(**1992**)

6.1 Introduction

6.1.1 Direct and Indirect Drug Design

Designing new drug candidates is a challenging task for which structural information on the target protein or protein-ligand complexes is of utmost importance. However, for many pharmaceutical interesting areas no-high-resolution structure of the target protein is available, due to the fact that the majority of these target proteins are membrane-bound proteins [26] for which structure determination is a very time-consuming and difficult task. Our knowledge about the three-dimensional structure of globular proteins has grown exponentially in the last decades. Combining this structural information with low-resolution structures of a target protein of interest with all the available biochemical experimental data for the target protein, enables researchers of creating a three-dimensional model structure of the protein. We pursued this approach for the human CCK-B receptor as described earlier in Chapter 4.
A model of the protein, especially of the ligand binding site, can be improved by including three-dimensional structure-activity relationship (3D QSAR) data of known ligands. These structure-activity data of ligands can be derived by extracting common structural features being prerequisite for strong binding [38] or by using methods such as CoMFA (Comparative Molecular Field Analysis) [210, 211] or CoMSIA (Comparative Molecular Similarity Index Analysis) [212, 213]. When these comparative studies are pursued without accounting for the protein, a so-called indirect drug design is employed. In contrast, a direct design approach utilises an experimentally derived or a modelled protein structure for the purpose of molecular design.
A pronounced disadvantage of indirect design methods refers to the alignment of the ligand set which can be performed in several cases only ambiguously. Slight changes in the conformation of the ligands or a modified selection of alignment features can have a huge influence on the resulting molecular superposition. A much better and less ambiguous alignment is defined using a protein binding site as restraint to define the ligands in their bound conformation and orientation. However, also in this case flexibility of the target protein, which can alter its conformation while accommodating individual ligands properly, can not be taken into account. Therefore the generated approximative binding sites do not necessarily resemble in detail the actual binding modes.
The strength of the in the following applied method in this study results from a combination of both design approaches. The ligand molecules are aligned to approximate the molecular description of the putative protein binding site. This is in contrast to other 3D-QSAR methods (e.g. CoMSIA [212, 213]) in which an overall superposition is achieved by matching the individual molecules sequentially onto a preselected reference molecule.
The unique feature of the applied DRAGHOME method is that the binding site of the protein and the different ligand molecules are translated into abstract molecular descriptions. These descriptions are, in analogy to the ligand alignment method SEAL (Steric and Electrostatic Alignment) [214] and the 3D-QSAR method CoMSIA, soft Gaussian functions, which approximate different molecular properties by functional property distributions in space.
In the following section, the theoretical foundations of this method are described.

6.1.2 The DRAGHOME Docking Method

The total molecular property field p($\overrightarrow{r}$) of a molecule is composed of single properties p_i located at $\overrightarrow{r_i}$ declining with a distance depending Gaussian type function [equation 6.1].

$$p(\overrightarrow{r}) = \sum_i p_i e^{-\alpha(\overrightarrow{r} - \overrightarrow{r_i})^2} \tag{6.1}$$

The most important advantage of these functions is that, due to their 'smooth' nature, any singularity at the origin of the individual properties is avoided. Therefore, no cut-off values are required. The property distributions describe the features of the molecules which matter for protein-ligand interactions; notably:

- volume occupancy in the binding site
- electrostatic properties
- H-bond donor
- H-bond acceptor
- lipophilicity

By the subsequent superpositions of the property distribution functions, the different molecules are compared by determining their relative spatial similarity in terms of overlap of their property distributions. Therefore, the similarity score S is proportional to the scalar product of the different property densities of the superimposed molecules [equation 6.2].

$$S = -\sum_{i \in p} \sum_{j \in L} p_i p_j e^{-\frac{1}{2}\alpha(\overrightarrow{r_i} - \overrightarrow{r_j})^2} \tag{6.2}$$

The weighted sum of the contributions [equation 6.3] of the different property types results in an overall description that is used to evaluate the placement of a ligand onto the binding site representation.

$$\begin{aligned} S_{total} = \; & w_{volume} S_{volume} + w_{electrostatic} S_{electrostatic} + w_{acceptor} S_{acceptor} + \\ & w_{donor} S_{donor} + w_{lipophilicity} S_{lipophilicity} \end{aligned} \tag{6.3}$$

After determination of the orientation of the ligand molecule and the most representative model of the protein, a ligand-receptor complex can be generated and relaxed in a subsequent MD-simulation. The resulting minimised structures of ligand molecules and protein can be used as subsequent input structures for a next alignment cycle. Following this approach, conformations of the ligand molecules and the protein can be refined and optimised in an iterative manner. A summary of the DRAGHOME method is given in the flowchart depicted in Figure 6.1.

6.1.2.1 Protein Binding Site Representation

A binding site is roughly assigned manually by selecting those protein atoms that fall into a volume occupied by a set of overlapping spheres with a 6 Å radius centred around the key atom of the side chain of an amino acid identified by mutagenesis supposedly as being involved in

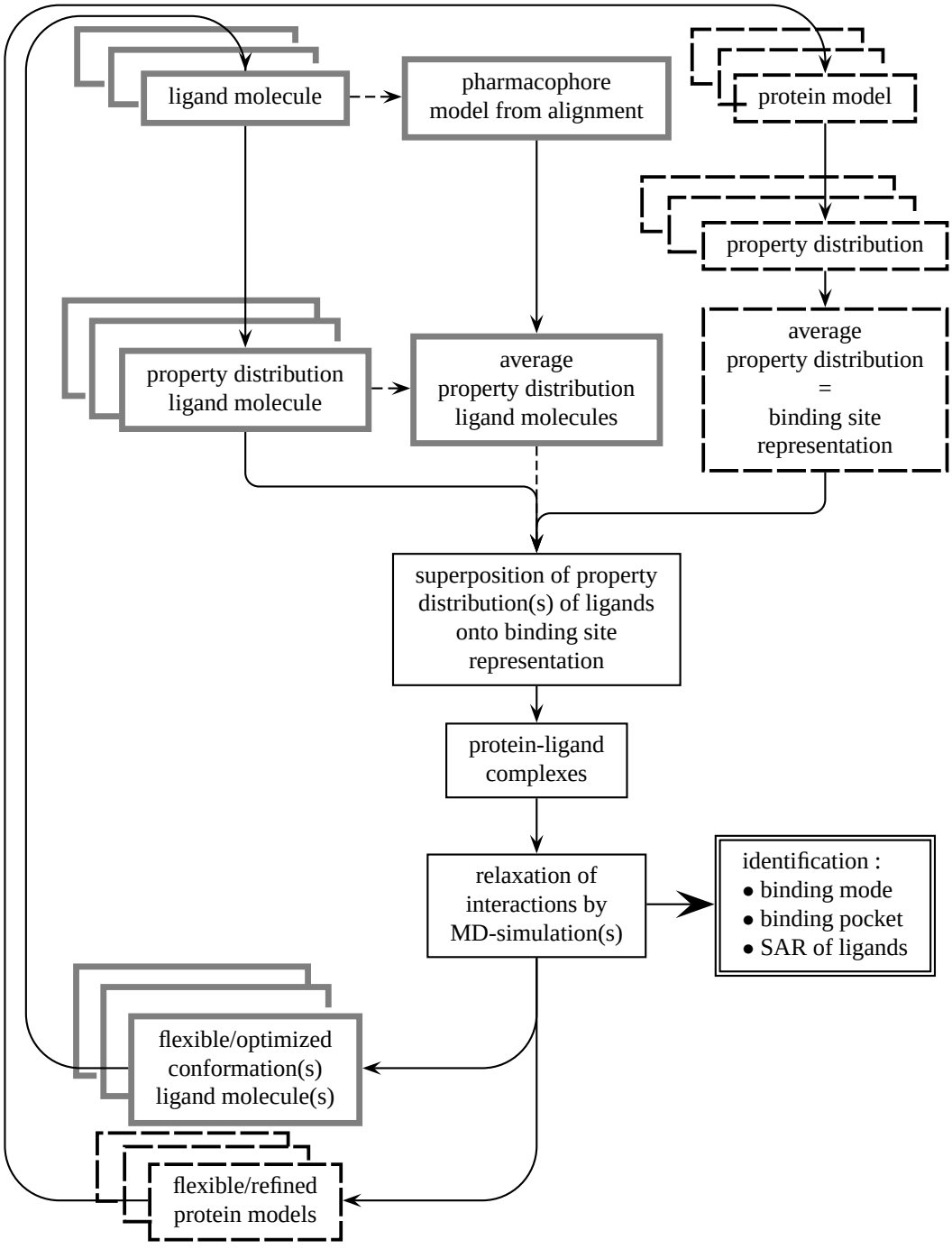


Figure 6.1: The docking procedure in this study mutually superimposes molecular property distributions of ligand molecules (gray-border boxes) and protein models (dashed boxes). Through an iterative refinement of both protein models and ligand conformations, the overlay is optimised and 3D-QSAR can be derived.

ligand binding. This area is then translated into a spatial description for the molecular properties of the protein using the above-mentioned Gaussian functions.
The donor, acceptor and lipophilicity property distributions are described by putative interaction sites generated with the program LUDI [215, 216]. LUDI maps the possible range of interactions by placing multiple interaction sites around the exposed functional groups of the protein residues according to rules derived from experimental information. Thereby, hydrogen-bond acceptors are represented by CO-groups, hydrogen-bond donors by NH-groups, and lipophilic sites by C-atoms, respectively. A weighting factor is applied according to the total number of sites placed around a particular functional group. The LUDI assignment of donor and acceptor functionalities is completed by searching over all protein N and O atoms, which could serve as putative hydrogen-bonding partners for a docked ligand, within an 5 Å sphere around the generated LUDI sites. Furthermore, the volume principially available to the ligand, is generated by merging all generated LUDI sites.
The volume occupied by the protein is approximated by Gaussian functions located at the positions of the protein atoms weighted by the radii of the respective atoms. In order to consider this protein volume in relation with the aforementioned distribution functions, the protein volume function is given an opposite sign. In addition, type-specific atom charges are assigned using Kollman charges [217] as implemented in TRIPOS' biopolymer force field.
In summary, the following molecular property functions are accounted for:

- ligand-accessible volume determined by LUDI
- ligand-inaccessible volume occupied by the protein atoms
- lipophilic sites of the accessible volume determined by LUDI
- putative H-bond donor partners determined by LUDI
- putative H-bond acceptor partners determined by LUDI
- charges attributed to the modelled protein atoms

The binding-site area is demarcated by applying soft boundaries, thus restricting the ligand to binding site.
The molecular property distribution functions assigned to the protein binding site can be determined for different models of the same protein. By averaging the assigned functions of different protein models a composite function is obtained, which reflects the flexibility and/or the uncertainty of the protein model.
For a given position $\overrightarrow{r}$ with property p_m in model m, the resulting average for N models is given by equations [6.4] and [6.5]:

$$p_{arithmetic}(\overrightarrow{r}) = \frac{1}{N} \sum_{m \in M} p_m(\overrightarrow{r}) \tag{6.4}$$

$$p_{geometric}(\overrightarrow{r}) = \left(\prod_{m \in M} p_m(\overrightarrow{r}) \right)^{\frac{1}{N}} \tag{6.5}$$

The resulting averaged property distribution is a new property distribution function. Arithmetic averaging can be regarded as logical OR. If there is a finite property value in at least one of the

models, the average will also possess a finite value at this point. On the contrary, the geometric averaging is a logical AND. Only if all models possess a finite property value at a given position the geometric average will also have a finite property value at his point. Since in this study a global model of the protein should be achieved, the arithmetic average procedure was used for averaging the distribution functions of the protein.

6.1.2.2 Ligand Representation

Similar to the representation of the repulsive protein volume, the volume of the ligand molecules is described by Gaussian functions associated to the atoms and weighted by the atom radii of the involved ligand atoms. The lipophilicity distribution function is assigned by using the parametrisation of Ghose and Crippen [218] for the ligand atoms similar to the determination of the volume distribution function.
The evaluation of the donor and acceptor distribution functions is based on the algorithms implemented into SEAL [214]. The putative hydrogen-bonding interaction sites around the ligand donor and acceptor groups are determined using a small set of predefined rules derived from small molecule crystal data. Similar to the protein donor and acceptor distribution functions, the ligand donor and acceptor distribution functions are weighted by the total number of sites assigned to a particular functional group of the ligand. In addition, charges are assigned using the Gasteiger-Hückel charges [219]
In summary, the following molecular property functions for the ligand are considered:

- ligand volume
- ligand lipophilicity
- putative H-bond donor partners
- putative H-bond acceptor partners
- charges attributed to the ligand atoms

Based on an alignment of the ligand molecules, an averaged property distribution for the aligned ligands can be determined. Usually the geometric average (equation [6.5]) and not the arithmetic average (equation [6.4]) is used, because the properties only present in *all* structures are then highlighted. The property distribution functions of such a ligand alignment can be considered as field-based pharmacophore of the ligands. (Figure 6.1)

6.1.2.3 Docking Ligands onto the Binding Site Representations

In order to reduce computational efforts, data reduction is applied. The initial distribution functions of volume and lipophilicity based on the atoms are approximated by distribution functions based on a reduced set of centres. After this data reduction, the superposition can be achieved straightforward.
The method for generating and optimising the ligand orientations during superpositioning was adapted from the ligand alignment program SEAL [214]. In the first step, a number of starting orientations is generated using a clique search algorithm. For each such starting orientation the similarity score (equations [6.2] and [6.3]) is calculated. The ligand orientation with respect

to the binding-site representation is optimised until the nearest local minimum of the similarity score is reached [220, 214]. Hereby, only the best-scored starting orientations beyond a pre-defined threshold were used. According to findings by Schafferhans et al., 50 starting orientations allow for a satisfactory sampling of orientation space [221].
As a result of the ligand superpositioning onto the binding-site representation, a number of possible solutions ranked according to their similarity score emerges for each ligand. Validation studies, preformed by Schafferhans [221] on crystallographically determined protein-ligand complexes, have shown that the experimentally determined structures are not always found as best solution. However, considering in addition the mutual similarity of the different ligand molecules onto the binding-site representation, the experimentally determined orientation of these ligands could be ranked as the best solution. This strategy to identify the best ligand-docking solutions was also applied in this study. For this purpose, a combination of visual inspection and computational analysis was performed.

6.1.3 Evaluation Docking Results

The difficult task to find the correct binding mode of a ligand with respect to its receptor is as old as the lock-and-key concept of Fischer [222]. It plays not only an important role in molecular docking, but also in the area of virtual database screening as well as for the scoring of *de novo* designed molecules. For these application a reliable, but nevertheless rapid method is needed to predict approximate binding affinities of a large and diverse set of ligands. Böhm [223] developed such a general-purpose empirical function to estimate the free energy of binding for a given protein-ligand complex of known 3D structure.

6.1.3.1 Binding Affinity

A quantitative knowledge of the binding process is essential to understand molecular recognition. This requires a detailed understanding of the physical forces involved in the interaction and a measure to what extent these forces contribute to the overall binding process. The receptor-ligand interactions are dependent on the concentration of the ligand and the protein, the type of solution and the salts in the solution. A simple equilibrium process is given in equation [6.6].

$$RL \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} R + L \tag{6.6}$$

In equation [6.6], R is the receptor, L the ligand, RL the receptor-ligand complex, k_{-1} the association rate constant and k_{+1} the dissociation rate for the equilibrium, respectively. The equilibrium constant for this reaction is given in equation [6.7].

$$K_d = \frac{[R][L]}{[RL]} \tag{6.7}$$

Through fundamental thermodynamic equations, the relation of the binding constant and the change in free energy (Gibbs energy) can be described as in equation [6.8].

$$\triangle G = \triangle H - T \triangle S \tag{6.8}$$

where $\triangle G$ is the change in free energy of a reaction, $\triangle H$ the change in enthalpy, $\triangle S$ the change in entropy, and T is the temperature of the system, respectively. The relationship between the free energy change and the dissociation of RL is given in equation [6.9]:

$$\triangle G = \triangle G^{\circ} + RT \ln K_d \tag{6.9}$$

At equilibrium under standard conditions $\triangle G = 0$:

$$\triangle G^{\circ} = -RT \ln K_d \tag{6.10}$$

and:

$$\triangle G^{\circ} = \triangle H^{\circ} - T \triangle S^{\circ} \tag{6.11}$$

$\triangle G^{\circ}$ is the standard free energy chance of the reaction when all of its reactants and products are in their standard states.
Combining the two equations [6.10] and [6.11] yields equation [6.12]:

$$\ln K_d = \frac{\triangle G^{\circ}}{-RT} = -\frac{\triangle H^{\circ}}{R}\left(\frac{1}{T}\right) + \frac{\triangle S^{\circ}}{T} \tag{6.12}$$

where R is the gas constant (8.3143 J $\text{mol}^{-1}\text{K}^{-1}$), T is the absolute temperature (in K) and $\triangle G^{\circ}$ is the free energy change (in J) associated with the reaction under standard conditions (all products/reactants present at 1M concentration, T = 298K and pressure 1 atm). This means that a 10-fold variation of K_d at $25^{\circ}C$ corresponds to a 5.7 kJmol^{-1} change in $\triangle G^{\circ}$, which is about the free energy of even a weak to medium hydrogen bond.
The binding constant of a potent drug is about $K_d = 10^{-9}$ $M = 1$ nM. At physiological temperature (310K) this relates to a change of -53.4 kJ/mol in free energy.

6.1.3.2 Scoring Function

In order to determine the binding constant, the difference between the free energies of the complexed (c) and the uncomplexed (u) states have to be calculated according to equation [6.13]:

$$\triangle G = \triangle G^{c} - \triangle G^{u} \tag{6.13}$$

Ajay and Murcko reviewed the complexities and the different approaches in calculating the free energy differences of ligand-protein complexes [224]. Apart from the molecular mechanics method to calculate the free energy difference, the "Master Equation" approach is another important and intensively used method. In this approach, the first step is to factorise the binding free energy into different components, such as:

- Overall loss in entropy due to association (loss in translational and rotational entropy of the rigid components (protein and ligand))
- Hydrophobic energy (entropy gain of water due to the binding of the ligand)
- Entropy loss of receptor and ligand due to immobilising rotational and translational degrees of freedom of ligand and receptor

- Interaction energy between ligand and receptor in which water molecules can play an important role. A major portion of this interaction energy is the electrostatic interaction.
- Changes in steric interaction energy on binding.
- Changes in conformational energy of the receptor and ligand upon binding.

An assumption which is made in approaches using this factorisation, is that the components of the free energy change can be calculated separately and that they are additive (equation [6.14]) [224].

$$\triangle G_{binding} = \triangle G_{motion} + \triangle G_{interaction} + \triangle G_{solvent} + \triangle G_{conformation} \tag{6.14}$$

An important simplification in the calculation of free energy is that the computed complex is expressed by only one unique structure. A more realistic picture has to consider an ensemble of structures which dynamically interconvert. However, it is assumed that the dynamic nature of the complex will not significantly affect the total estimate of binding energy difference. The components and accuracy of scoring functions have been reviewed [225, 226, 227, 228, 229].

6.1.3.3 Pharmacophore Determination

Apart from binding affinity prediction using a scoring function, pharmacophore determination is a further method to evaluate docking results of a series of ligands.
A pharmacophore is a 3D-model describing in a semi-quantitative manner the requirements imposed on to a ligand compound in order to experience a certain binding affinity towards the receptor protein. The quantitative correlation of molecular structure with binding affinity of a ligand towards its receptor is the goal of a quantitative structure-activity relationship (QSAR) study. Subsequently, the derived correlation model can be used to predict the binding affinity for novel compounds. A variety of methods are available, such as the classical QSAR methods, e.g., the Hansch-Analyse or the Free-Wilson method, but also 3D-QSAR methods, e.g., CoMFA [210, 211], CoMSIA [212, 213] and SEAL-field comparisons [214, 230]. In this study, we have applied the three mentioned 3D-QSAR methods.
The 3D-QSAR methods CoMFA [210, 211] and CoMSIA [212, 213] are grid-based approaches to derive an active-site model for the receptor in terms of molecular electrostatic, steric and hydrophobicity fields. In the CoMFA analysis, a sufficiently large box containing a grid with a certain grid spacing is positioned around the aligned molecules. Different atomic probes, e.g., a carbon atom, a positively or negatively charged atom, a hydrogen bond acceptor or donor, or a lipophilic probe are used to calculate field values on each grid point for each molecule. These field values are the energy values, the probe would experience at the corresponding grid intersections when it would interact with a particular molecule. A PLS analysis is used for correlating the biological activities with the field values. Normally, cross-validation is used to check the predictive power of the derived model. [211]
Within CoMSIA, SEAL similarity fields [212, 213] are calculated at the grid points. These similarities are not a direct measure of similarity between all mutual pairs of molecules, but they are indirectly evaluated via the similarity of each molecule in the data set with a common probe atom placed at the grid points. The advantage of CoMSIA is the smooth nature of fields with a Gaussian type distance dependence. This is in contrast to the CoMFA fields which

have singularities at the atomic positions since Coulomb or Lennard-Jones potentials are used [212, 213].

The results of the CoMFA/CoMSIA analysis are usually evaluated in terms of contribution maps. The contour maps display favourable and unfavourable steric regions, favourable and unfavourable regions for electropositive or electronegative substituents, or favourable and unfavourable regions for lipophilic or hydrophilic substituents in certain positions of the molecules. [210, 211, 212, 213]

It is recommended that the set of molecules is sufficiently large and diverse in structural and affinity terms. Furthermore, the prediction of the binding affinity of novel compounds requires that they resemble the molecules in the training set, thus the method can only interpolate, not extrapolate. Underlying assumptions of the method are: (i) all molecules follow the same mode of action, (ii) occupy the same receptor binding site and (iii) adopt a comparable binding mode. 3D-QSAR methods are only reliable for relative comparisons of related structures. The approaches can only interpolate, attempts to extrapolate beyond the information comprised in the data used for training cannot provide reasonable results. Furthermore, the results of a CoMFA or CoMSIA analysis are not only depending on the training dataset, but also on the alignment of the structures and the conformations of the structures of the training set. The alignment should represent as closely as possible the relative geometry of the compounds when accommodated in the binding site of the receptor protein. In order to prove the validity of the derived CoMFA/CoMSIA-model, the consideration of an independent test set is mandatory [211].

A possible method to generate a reliable alignment of a set of molecules is implemented into the SEAL program [214, 230]. The alignment condition of SEAL is based on the mutual similarity indices pairwise calculated between all atoms of the molecules being compared. The program tries to align two molecules rigidly so that similar properties are located close to each other in both molecules. A molecule is described in terms of set of spatial distributions of charge, density (volume), lipophilicity and hydrogen-bond features, whereby every atom is represented by a Gaussian distribution endowed with a property vector. A molecule corresponds to the sum of such distributions. The SEAL program has been extended and improved by Klebe et al. [230]. The program TORSEAL [230] uses the alignment condition of SEAL, however it applies a geometry optimisation in torsion angle space considering van der Waals interactions together with empirically derived potentials. The later potentials are based on statistically evaluated torsional fragments retrieved form the Cambridge Structural Database. In the optimisation step, the initially generated geometries are minimised to remove unfavourable steric contacts between adjacent atoms in a molecule, simultaneously keeping the torsion angles as close as possible to the most frequently occurring values in experimental structures. TORSEAL fits flexibly a set of molecules onto a rigid reference simultaneously considering atomic properties such as hydrogen-bond acceptor and donor abilities, the charge distribution and lipophilicity for similarity fields in the alignment procedure [230].

However, an alignment can also be obtained as the result of a docking procedure. Both kind of alignment procedures were used in this study.

The property fields of the SEAL method can also be applied in the development of a pharmacophore. For all aligned molecules, property fields are calculated and these property fields are geometrically (the AND-method) or arithmetically (the OR-method) averaged. The new fields are then approximated by new sums of Gaussian distributions and these can be visualised by contour maps representing the features of a field-based pharmacophore.

6.2 Method

All calculations, analyses and molecular simulations were carried out on Silicon Graphics O2 workstations and on a Silicon Graphics Origin parallel computer equipped with 20 R12000 processors. The interactive model building and graphical analysis were mostly done using the programs included in the molecular modelling software package INSIGHT II (version 98.0) from Biosym Technologies (San Diego, CA). Molecular dynamics simulations and energy minimisations were performed using the consistent valence force field (cvff) without Morse potential or cross terms as implemented in the DISCOVER simulation package (version 2.98). All the comparative molecular field evaluations and some of the interactive modelling for the DRAGHOME docking procedures were performed using SYBYL version 6.5 and 6.6.

6.2.1 The Protein Model

The receptor protein of interest in this study is the human CCK-B receptor. The construction of a 3D-model for this receptor was described in Chapter 4. The different conformations needed for the docking procedure with the DRAGHOME program were derived from a molecular dynamics simulation, which was described in Chapter 4. The 750 different conformations of the last 750 ps (301-1050 ps) of the molecular dynamics simulation are considered as representatives covering the conformational space accessible to the protein. The program NMRCLUST [231] was used to determine the most representative models out of these 750 conformations. This program superimposes the different conformations and subsequently clusters the superimposed conformations with respect to the root mean square deviation (RMSD) of several selected atoms. In order to be able to cluster the structures and to derive the most representative conformations of these clusters, reference atoms have to be assigned. The atoms used for superpositioning were the backbone heavy atoms of the determined transmembrane domain sequences (refer to Chapter 3: residues I:33-59; II:64-90; III:103-137; IV:145-169; V:197-226; VI:241-270 and VII:287-310 of the human CCK-B receptor). As clustering atoms, the keys atoms of the amino acid residues were selected, which were determined in several mutagenesis experiments [147, 232, 233] to be important for antagonist binding (see Table 6.1). The applied pre-set threshold for the clustering cut-off was set to 1.25 Å in order to obtain a reasonable number of clusters.
To estimate the conformational behaviour of these 14 residues, their χ_1 and χ_2 sidechain torsions were analysed.
The obtained representative conformations of the receptor were used in the subsequent procedure of building the protein binding site representation (6 Å around each key atom of Table 6.1) with the DRAGHOME program.

6.2.2 The Ligands

In Chapter 5 an overview of the CCK-B antagonists was given. In Table 6.2, a selection of these compounds is listed which were investigated in this docking study. These compounds all possess a rather rigid structure. Furthermore, they exhibit a large structural variation with respect to their attached substituents and a large spread in biological activity.
A search in the Cambridge Structural Database (CSD) was performed to retrieve compounds with close structural similarity to our ligands in order to suggest a plausible starting conforma-

residue no. in human CCK-B	amino acid	key atom	ref. mut. data
57	Arg	NE	[147]
100	Asp	OD1	[150]
111	Thr	OG	[147]
131	Ser	OG	[147]
219	Ser	OG	[147, 149]
227	Phe	CZ	[232]
342	Phe	CZ	[232]
346	Trp	NE1	[149, 232]
349	Val	CG1	[147, 233, 149]
350	Tyr	OH	[149]
353	Asn	OD1	[149, 234]
356	Arg	NE	[147]
376	His	ND1	[147]
379	Ser	OG	[149]

Table 6.1: Key atoms of amino acid residues that were identified by mutagenesis as being involved in antagonist binding.

tion. These conformations were then used as basis for the modelling of the remaining ligands.

As an initial test for the general applicability of the docking approach, a first rigid docking trial was performed for the 41 ligands of Table 6.2, using only one conformation for 34 ligands and two conformations (pseudo-axial and pseudo-equatorial, see Figure 5.14 Chapter 5 [208]) for 7 ligands. These conformations were obtained by a mutual alignment using TORSEAL [230]. For this purpose, compound **89** was used as the reference molecule, since it is the most rigid entry in the data set. Since TORSEAL performs only a local optimisation into the next local minimum of the alignment space a reasonable preorientation of the considered molecules has to be ássumed. For the benzodiazepine derivatives it is clear that the benzodiazepine moieties should be aligned. However, the superpositioning of the quinazolinone derivatives and the diphenylpyrazolidinone derivatives onto the benzodiazepine derivatives is difficult, since they do not share obvious structural similarity with the benzodiazepine derivatives.
After the initial rigid docking trial the following points were considered:

- One of the strengths of the docking approach of DRAGHOME is the independent docking of the ligands within the representation of the binding-site. However, by assuming an initial mutual prealignment, there will be some dependence of the generated conformations of the docked ligands with respect to the conformation of the reference molecule.
- Although, the alignment obtained by TORSEAL indicates that the ligands have similar property fields, the initial mutual alignment was not reproduced in the first rigid docking trial. This might indicate that several ligands as, e.g., the quinazolinone derivatives, might adopt a slightly different binding mode and/or conformation than was suggested by the mutual alignment.
- The quinazolinone derivatives and the diphenylpyrazolidinone derivatives have possibly a slightly different binding mode in respect to compound **89**. This could explain the diffi-

compound nr., firm name	CCK-B affinity [nM]	exp. △G [kJ/mol]	axial and/or equit.
44 LY-247348	26	-44.372	
48	585	-36.462	
49	137	-40.150	
50	14	-45.945	
51	108	-40.754	
52	29	-44.095	
53	1.0	-52.650	
54	11	-46.558	
55	97	-41.027	
56	63	-42.124	
57 LY 288512	370	-37.626	
58 LY 288513	19	-45.169	
60 LY 294920	1900	-33.469	
61 LY 294919	550	-36.619	
31 L-364,718	270	-38.427	E + A
66 "L-364,718-R"	3700	-31.776	E
71"L-365-260-S"	151	-39.903	E + A
72 L-365,260	2.0	-50.889	E
73 L-740,093	0.1	-58.500	E
73S "L-740,093-S"	N.T.[1]		E
74 L-708,474	0.28	-55.884	E
75 YM-022	0.11	-58.258	E + A
75S "YM-022-S"	N.T.[1]		E + A
76 L-368,935	0.14	-57.645	E
77 L-738,425	0.11	-58.258	E + A
78	0.58	-54.034	E
79	7.5	-47.531	E
80	1.1	-52.408	E
81 L-736,380	0.054	-60.065	E
82	137	-40.150	E
83	1.5	-51.620	E
84	54	-42.515	E
85	18	-45.307	E
87	480	-36.965	E
88	23	-44.684	E
89	8	-47.367	E + A
90	2	-50.889	E
92	84	-41.393	E
97	1	-52.650	E
98	48	-42.815	E
100	26	-44.372	E + A

[1] N.T. = not tested.

Table 6.2: The compounds of Chapter 5 used in this docking study. The last coulom describes for the benzodiazepine derivatives which orientation of the 3-amide sidechain was used, namely the pseudo-axial (A) and/or the pseudo-equatorial (E) orientation (see also Figure 5.14).

culties in the alignment of the quinazolinone derivatives and the diphenylpyrazolidinone derivatives to compound **89** using TORSEAL.

- The initially assumed prealignment for TORSEAL biases the conformations of the ligands used for rigid docking.

- A more realistic approach would be, not only to account for receptor flexibility and receptor conformational uncertainty by using several receptor conformations (see section 6.2.1), but also to account for ligand flexibility and a possible multiplicity of ligand binding modes by using multiple conformations of each ligand.

Therefore, it would be preferable to dock in a second docking trial multiple conformations of a ligand and select after the docking step, the best docked conformation of this ligand based on the criteria imposed by the binding site representation. Schafferhans et al. [221] also tackled the problem of accounting for ligand flexibility by using pre-calculated multiple conformations for each ligand.
In order to examine the conformational freedom of the ligands and to obtain the possible conformations of a ligand, Monte Carlo-based conformational analyses (MOCCA) [235] were performed for all the ligands. The parameters were set to the values given in Table 6.3. In this approach all rotatable bonds are allowed to move. However, by allowing only the torsion angles to rotate, the so-called "flipped" conformation of the benzodiazepine derivatives (the 3-amide side chain in the pseudo-axial position instead of the pseudo-equatorial position, see Figure 5.14, Chapter 5 [208]) could not be obtained in a MOCCA analysis [235]. Consequently, for some benzodiazepine derivatives MOCCA analyses were done with a "flipped" benzodiazepine ring conformation (see Table 6.2).

parameter	value
force field	cvff
dielectric constant	80
start temperature of the simulation	10000 K
end temperature of the simulation	300 K
number of Monte Carlo steps	1000
start angle change	180°
exponent for the torsion angle scaling	-0.002303
scaling factor for hydrogen-bonds	0.70
scaling factor for the chirality energy	1.0
standard randomvalue	4711
number of structures	900

Table 6.3: The parameters of the MOCCA program used for the calculations.

6.2.3 Docking of the Ligands

As mentioned before, an initial docking trial was performed in order to test the applicability of the docking approach. In this initial trial a rigid docking was performed for the 41 structures of Table 6.2, using only one conformation for 34 ligands and two conformations for 7 ligands.

These conformations were obtained by mutual alignment using TORSEAL. The docking of these 48 structures was done interactively and the results were analysed by visual inspection of the obtained solutions for every ligand.

A flexible docking was performed in the second docking trial by using pre-calculated multiple conformations for each ligand generated by a MOCCA analysis [235]. The docking procedure and the selection of the best solutions for each ligand in this docking trial were automated by computer scripts, since 900 conformations for each structure had to be docked in the binding site representation.

For each docking solution, DRAGHOME gives a normalised total score, a hydrophobic, a steric, an electrostatic, a donor, and an acceptor score, respectively. The total score is a weighted sum of the five different scores (equation [6.3]). This total scores was optimised by reproducing best the ligand binding modes for a set of crystallographically determined thrombin-inhibitor complexes. For this test set the electrostatic property contributed very little to the total score [221]. Possibly, the electrostatic property displays redundant information, since it is located mainly in the same regions as the acceptor and donor properties [221]. It is therefore likely that the considered physico-chemical properties are not linearly independent and in the present study it is assumed that excluding the electrostatic property will not affect the results significantly.

Furthermore, Schafferhans et al. [221] mention that the total score is strongly dominated by the steric property. The normalised contributions of the five properties is 98.0, 1.9, 2.9, -2.9 and <0.1 for steric, acceptor, donor, lipophilicity and electrostatic properties, respectively. However, the total variance of the properties is more important than their contribution, since orientations are optimised and ranked according to their scoring differences. If the contribution of one property is almost constant, it cannot render prominent one particular orientation. Although the steric contribution to the score variance is still larger than the contributions of the other properties, it is less dominant than the steric contribution to the total score. The contributions to the variance are for the steric, acceptor, donor, lipophilic and electrostatic properties, 7.4, 0.8, 1.6, 1.0, and 0.1, respectively [221].

In the initial visual inspection of the docking results obtained in the first initial rigid docking trial with 48 structures, the domination of the steric property was clearly observed. The steric property dominated the total score rather strongly, thus the ranking of the docking solutions was mainly based on the steric contribution and the contributions of the other properties are of very little importance. It can be hypothesised that the weighted sum, optimised by using thrombin-inhibitor complexes, is not necessarily representative for the application to other proteins. Therefore, other criteria than the total score given by DRAGHOME were applied to select the best docking solutions out of the different solutions produced by DRAGHOME.

The first method is based on the idea that a solution with high rankings in all four different properties will be a good solution. The variance within a property, represented by the rankings, is in this selection method more important then the quantitative score of a property. For each docking performed by DRAGHOME, the scores of each property were ranked. The rankings were multiplied with an exponential function in order to favor the high rankings more than the low rankings. The resulting weighted ranking number of the four different properties were averaged to a total weighted ranking number following a consensus scoring philosophy. The best solution is then the solution with the lowest total weighted ranking number. This is the solution on average ranked the highest in all the four scores.

In the second flexible docking trial, for every docked conformation of a ligand the best two solutions were selected with the lowest total weighted ranking number. These selected solutions (1800 for every ligand) were for each ligand again ranked to each other using the above described ranking procedure. The 10 solutions with the lowest total weighted ranking number were selected for visual inspection.
Schafferhans et al. [221] evaluated a strategy to identify the best solutions on the basis of the spatial similarity of various solutions generated for the different ligands for internal similarity consistency, even if they are not scored on the first rank by DRAGHOME. This strategy is of course only valid as long as the ligands bind to the same binding site. A similar strategy can also be applied in this study.
On basis of the spatial similarity strategy of Schafferhans et al. [221] and the visually inspected results an extra selection rule was defined. The best 10 solutions of compound **76** (L-368,935), a ligand with a high affinity for the CCK-B receptor, were all in a very similar position. Their urea-group, an essential group for CCK-B antagonists (see Chapter 5), was positioned in close proximity to the amino-acid residue Ser379. Also many solutions of the other benzodiazepine ligands are found in a similar orientation. Therefore, it was hypothesised that an acceptable solution should have its urea group, or at least a similar group, in an area of 3 Å around the urea group of compound **76** (L-368,935). This selection rule will be called in the following the "urea-group"-rule. For the quinazolinone and the diphenylpyrazolidinone ligands both O- and/or S-atoms were applied subsequently in this "urea-group"-rule.
The number of solutions from the top 100 which follow this rule were counted and could be an indication for the quality of the docking in analogy to a selection criterion of the docking approach of Baxter et al. [236]. The solution which followed the "urea-group"-rule and had the highest rank/the lowest total weighted ranking number was selected as the best solution of the ligand. In an ideal situation it could be found on the first rank.

However, the selection of the best solutions on basis of the average of the weighted rankings of the four properties combined with the "urea-group"-rule is not a validated selection criterion, since this strategy is not tested on ligand-receptor complexes for which the experimental binding affinity and binding geometry are known. Furthermore, the results are depending on the applied weighting functions and the "urea-group"-rule is only based on the hypothesis that the amino-acid residue Ser379 interacts with an O- and/or S-atom of the ligand. Mutagenesis data suggest that Ser379 plays a role in ligand binding, but its specific role in the binding modus of a ligand is not yet known. Therefore, a second selection strategy was applied.
This selection strategy is not based on such hypothetical rules. It takes advantage of the large number of solutions created by DRAGHOME by docking a large number of different conformations for a single ligand (900 conformations for each ligand). All solutions created by DRAGHOME for one ligand were clustered on the basis of the root mean square deviation (RMSD) of their heavy atoms. This grouping method is not a stringent clustering that can be performed with the program NMRCLUST [231], since this kind of clustering of all solutions obtained by DRAGHOME for one ligand would be computationally too intensive. The applied method calculates the RMSD between the considered solution and the first member of a group. When this RMSD is smaller than a given RMSD threshold, the solution is added to this group. However, when the solution could be assigned on basis of the threshold to several groups, it is merged to the group with the smallest RMSD value. If all calculated RMSD values are greater than the pre-defined RMSD threshold, the solution underconsideration forms a new

group. Thus, at the end, all members of a group have a RMSD value to the first member of the group smaller than the RMSD threshold, but not necessarily to each member of the group. Furthermore, the RMSD value between a member of one group and a member of another group could be lower than the RMSD threshold, but the RMSD value between the first member of a group and the first member of any other group is always greater than the RMSD threshold. Schafferhans mentioned in her thesis [237] that solutions with a RMSD smaller than 2 Å can be considered as solutions with similar orientations within the binding site representation. Therefore, a RMSD threshold of 2 Å was applied in the final grouping in order to obtain groups of solutions whereby the different groups represent different orientations and the members of one group have similar orientations. However, since the computing time is highly depending on the number of groups formed, first a preclustering was performed with a threshold of 7 Å. The most populated groups (the number of solutions) with the highest number of hits (every solution can be obtained several times in one docking trial by DRAGHOME and the summation of these countings over all members of a group is the amont of hits) and a low average of the total scores of all the members were selected for the final grouping. In cases were these three criteria do not hold simultaneously, the criterion considering the highest number of members is dominant, whereby the number of hits has to be at least twice the number of members of the group and the average of the total scores of the members has to be in the top 10 of lowest averages of the total scores. This selection of the best groups is done by manual intervention. Therefore, the selection is not totally unbiased, but the selection was done with the utmost care and consistency.
The solutions of the selected groups were clustered in a second attempt, but now using a RMSD threshold of 2 Å. The best group with the best solutions of DRAGHOME was selected by the same selection strategy as was applied in the preclustering. However, a spatial similarity criterion based on the strategy of Schafferhans et al. [221] was also taken into account during the visual inspection of the results.

6.2.4 Energy Minimisation and Molecular Dynamics Simulations

The binding-site representation calculated by DRAGHOME is an average representation. Consequently, complexes modelled by applying the DRAGHOME-derived positions of the docked ligands with one of the representative models of the receptor are not of refined geometry, since they still could be affected by unfavourable contacts between receptor and ligand. Therefore, a scoring function to rank the docking solutions is not yet applicable. The structure of such a complex has to be relaxed following an energy minimisation and/or a molecular dynamics simulation. The details of the protocols of these calculations are summarised in Tables 6.4 and 6.5.
For every ligand five complexes were modelled by placing into five different receptor models the best DRAGHOME solution following the above-described selection strategy including the weighted rankings of the four properties and the “urea-group”-rule. The five receptors models were the representative conformations of the top five clusters of the receptor conformations of the molecular dynamics run of Chapter 4 (see Table 6.6). The computationally more extensive protocol of Table 6.5 was used only for selected complexes. The best DRAGHOME solution of the compounds **50**, **53**, **57**, **66**, **73**, **75**, **80**, **81**, **84**, and **89** were placed into the binding site of the representative conformation of cluster 4, frame 393 of the molecular dynamics run of Chapter 4, since it is the most representative conformation of the 14 clusters of the receptor.

	method	number of step	force kcal/Å	fix/tethering
1.	template force, steepest descents	1000	100	fix backbone protein tethering urea-group + side chains protein force on urea-group torsions
2.	template force steepest descents	500	100	fix backbone protein tethering urea-group + side chains protein force on urea-group torsions
3.	minimize steepest descents	500	-	fix backbone protein force on urea-group torsions
4.	minimize conjugate gradients	1000	-	fix backbone protein force on urea-group torsions

Table 6.4: Energy minimisation protocol of the protein-ligand complexes.

	method	number of steps	force [kcal/Å] or temp. [K]	fix/tethering
1.	template force steepest descents	2000	1000	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
2.	template force conjugate gradients	2000	1000	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
3.	template force steepest descents	2000	100	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
4.	template force conjugate gradients	2000	100	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
5.	template force steepest descents	2000	10	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
6.	template force conjugate gradients	2000	10	fix backbone TMDs protein tethering side chains TMDs protein force on urea-group torsions
7.	minimize steepest descents	5000	-	fix backbone TMDs protein force on urea-group torsions
8.	minimize conjugate gradients	5000	-	fix backbone TMDs protein force on urea-group torsions
9.	initialize dynamics	2000	10	force on urea-group torsions
10.	resume dynamics	48000	10	force on urea-group torsions
11.	resume dynamics	100000	150	force on urea-group torsions
12.	resume dynamics	75000	300	force on urea-group torsions
13.	resume dynamics	25000	150	force on urea-group torsions
14.	resume dynamics	25000	10	force on urea-group torsions

Table 6.5: Energy minimisation and molecular dynamics simulation protocol of the protein-ligand complexes.

6.2.5 CoMSIA Analyses

For the CoMSIA analysis the standard SYBYL parameters were applied. The grid spacing of the lattice was 1 Å and by changing the grid box coordinates manually consistent grids were used in both analyses.
The statistical significance of the models were checked by leave-one-out cross validations with the SAMPLS method. The optimal number of components was determined by selecting the smallest s_{PRESS} value which usually corresponds to the highest q^2 value. They were computed as defined in SYBYL. The same number of components was subsequently used to derive the final QSAR models.
By scrambling the biological data and repeating the model derivation process, the consistency of the derived models were checked. Negative values for q^2 in the PLS analysis are then expected.
These CoMSIA analyses were done for two alignments of the ligands. The first alignment was obtained by using the best solution of the second flexible docking trial with DRAGHOME selected with the strategy on the basis of the average of the weighted rankings of the four properties in combination with the "urea-group"-rule. The second alignment was obtained by visually selecting for each ligand of the first alignment a representative conformation of the five minimised conformations of each ligand (see also section 6.2.4).

6.2.6 Docking of CCK-4 and CCK-8

For reasons of qualitative comparison between the binding of antagonists and endogenous ligands, an additional docking procedure was carried out for two peptides, notably CCK-4 and CCK-8. For this purpose, the same procedure was followed as described for the antagonists. Only within the MOCCA [235] calculations distance restraints were applied in order to avoid artificially contracted conformations, since the peptide binds in an extended conformation to the receptor [120]. Again, 900 conformers were created for each ligand which were all subsequently docked in the binding site representation created by DRAGHOME using the same procedures as were applied for the antagonists.

6.3 Results and Discussion

6.3.1 Protein Receptor Conformations

The 750 conformations of the 3D-model of the human CCK-B receptor described in Chapter 4 were clustered into 14 clusters by NMRCLUST [231]. The details of the RMS distribution of the clusters and the number of members are given in Table 6.6. The representatives of these 14 clusters were used in the subsequent docking procedures.
In order to determine the flexibility of the side chains of the 14 amino acid residues identified by mutagenesis as important for ligand binding (Table 6.1), their χ_1 and χ_2 torsion angles were evaluated (except for Val349, which has no χ_2 torsion angle). Graphs of the torsion angles χ_1 versus χ_2 are depicted in Figure 6.2 for the 13 amino acid residues. It can be seen that Arg57, Phe227, Phe342, Trp346, Tyr350 and His376 adopt a preferred sidechain conformation, Thr111, Asn353, Arg356 and Ser 379 display a moderate distribution, while Asp100, Ser131 and Ser219 show almost unrestricted conformational variability around their χ_2 torsion. Only

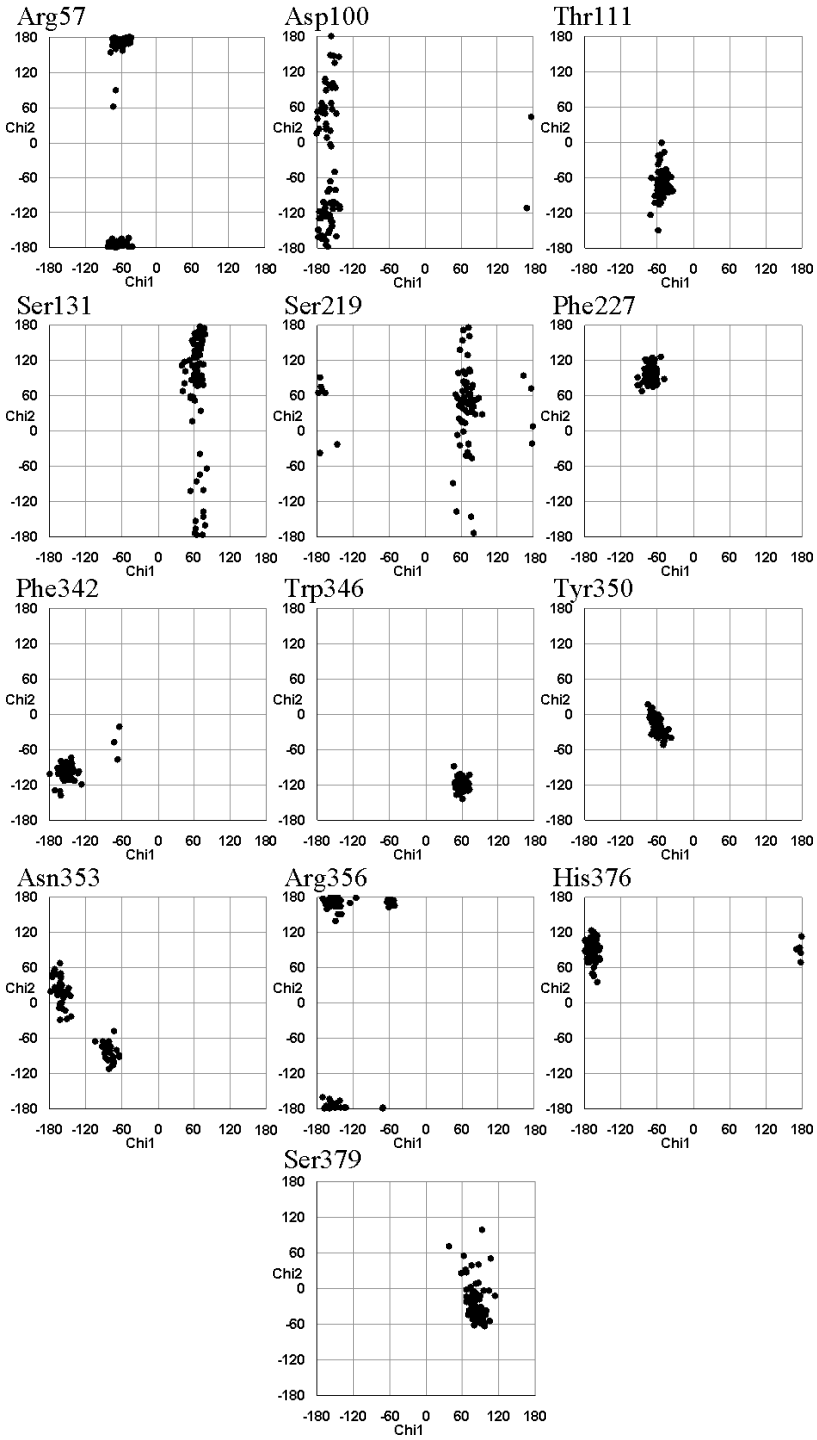


Figure 6.2: χ_1 and χ_2 torsion distribution of 14 representative receptor conformations of amino acid residues identified bu mutagenesis as being crucial for antagonist binding.

cluster	number of members within cluster	spread in RMS (Å)	frame number of representative conformation
1	187	1.116	1040
2	109	1.139	687
3	94	1.074	753
4	73	1.036	393
5	70	1.106	565
6	69	1.059	618
7	38	1.078	663
8	30	0.924	819
9	26	1.112	328
10	17	0.967	321
11	11	1.022	343
12	9	1.051	526
13	8	0.799	811
14	7	0.914	306
outliers	2	-	
total:	750		

Table 6.6: Size, RMS-spread, and frame number of the reprensative conformation of the clusters created by NMRCLUST [231].

Arg57, Ser219, and His376 expose their side chains in or in close proximity to the water compartment, while the other side chains are all buried within the protein as discussed in Chapter 4.

6.3.2 First Docking Attempts

Before we applied the DRAGHOME docking method in this study, the docking programs FlexX [238, 239] and FlexiDock (a module in SYBYL) were tested. However, the docking results of FlexX were not in agreement with any experimental evidence, as, e.g., the mutagenesis data (Table 6.1). The program was only able to identify binding modes at the outside of the receptor, i.e. the extracellular loop region. The results of FlexiDock were better, but they were highly depending on the start conformation of the protein-ligand complex, thus giving no convergent results. Thus both docking routines were not further applied in this study.

6.3.3 Docking with DRAGHOME

6.3.3.1 Representation of the Binding Site

The DRAGHOME program calculates the representation of the binding site based on five molecular property functions:

- volume accessible to the ligands
- volume inaccessible for the ligands

- lipophilic sites inside the accessible volume
- putative H-bond donor partners
- putative H-bond acceptor partners

These five calculated molecular property functions of the binding site can be depicted in terms of contour maps as shown in Figures 6.3, 6.4 and 6.5, respectively. The extracellular and intracellular membrane planes are approximately at contour-map-levels of +6 Å and -26 Å, respectively. These contour levels correspond to the dimensions of the box around the representation, it is not a measure of the membrane height.
From the 12 volume contour maps it can clearly be seen that cavities exists within the upper part of the transmembrane domain of the receptor. From contour map +8 to 0 Å and from -4 to -12 Å cavities between the helices 3, 4, 5, 6 and 7 can be identified. Between the helices 1, 2, 3 and 7 a cavity stretches from contour map 0 to -4 Å. The first two cavities are located approximately at the position of the binding site of the mono-amine receptors (see Figure 2.3) [240, 118]. The third cavity is located near the important residue His376.
As can be expected, donor and acceptor sites are found in spatial proximity of the backbone atoms of the receptor protein. However, in several contour maps the distribution functions of putative donor and acceptor partners possess high values at sites between the helices. A dense donor functionality can be found in the contour map at +8 Å in the upper part of the first cavity in proximity to helix 3 and in proximity to the amino acid residues His207, Arg208, Arg356 and Tyr192. Further, in map +2 Å between the helices 3 and 7 there is a high density of donor functionality, probably caused by Tyr189 at helix 4, Met134 at helix 3, His376 at helix 7 and several backbone atoms, respectively. The area between helices 3 and 7 is in numerous contour maps assigned to a high density of donor and/or acceptor functionality. The contour maps between 0 and -6 Å of the acceptor functionality and the contour maps between -4 and -8 Å of the donor functionality show enhanced densities. Most likely, both densities orginate from two amino-acid residues which possess dual character being either donor and/or acceptor. These two residues are Ser137 and Ser379 at helix 3 and helix 7, respectively. The later residue has already been rationalised as important for antagonist binding through mutagenesis studies [147] and Ser137 is located one helical turn below the position of the key aspartate residue of the mono-amine receptors [79], supporting the hypothesis of a similar location of binding sites for the small ligands within different GPCRs [32, 28].
In the upper part of the second cavity, a high density of acceptor functionality is found which might result from the presence of Asn353 at helix 6. Also this residue has been identified as being involved in ligand binding. However, not only these enhanced putative donor and acceptor functionalities at several locations in the receptor are interesting, furthermore the lack of comparable profiles between helices 3, 4, 5, 6, and 7 in the acceptor distribution maps between -4 and +14 Å and in the donor distribution maps 0, -6 and -8 Å are remarkable. When this indication is combined with the observation that in exactly this area of the contour maps a strong contribution of lipophilic potential is found, it can be hypothesised that the second cavity which is located in the same area is of rather hydrophobic nature with only a few possibilities for hydrogen bonding. In contrast, the third cavity has almost no hydrophobic properties as shown in the contour maps from 0 to -4 Å.
It should be emphasised that the contour maps shown in Figures 6.3 to 6.5 are not computed for rigid entities, they represent an averaged property distributions across 14 representative

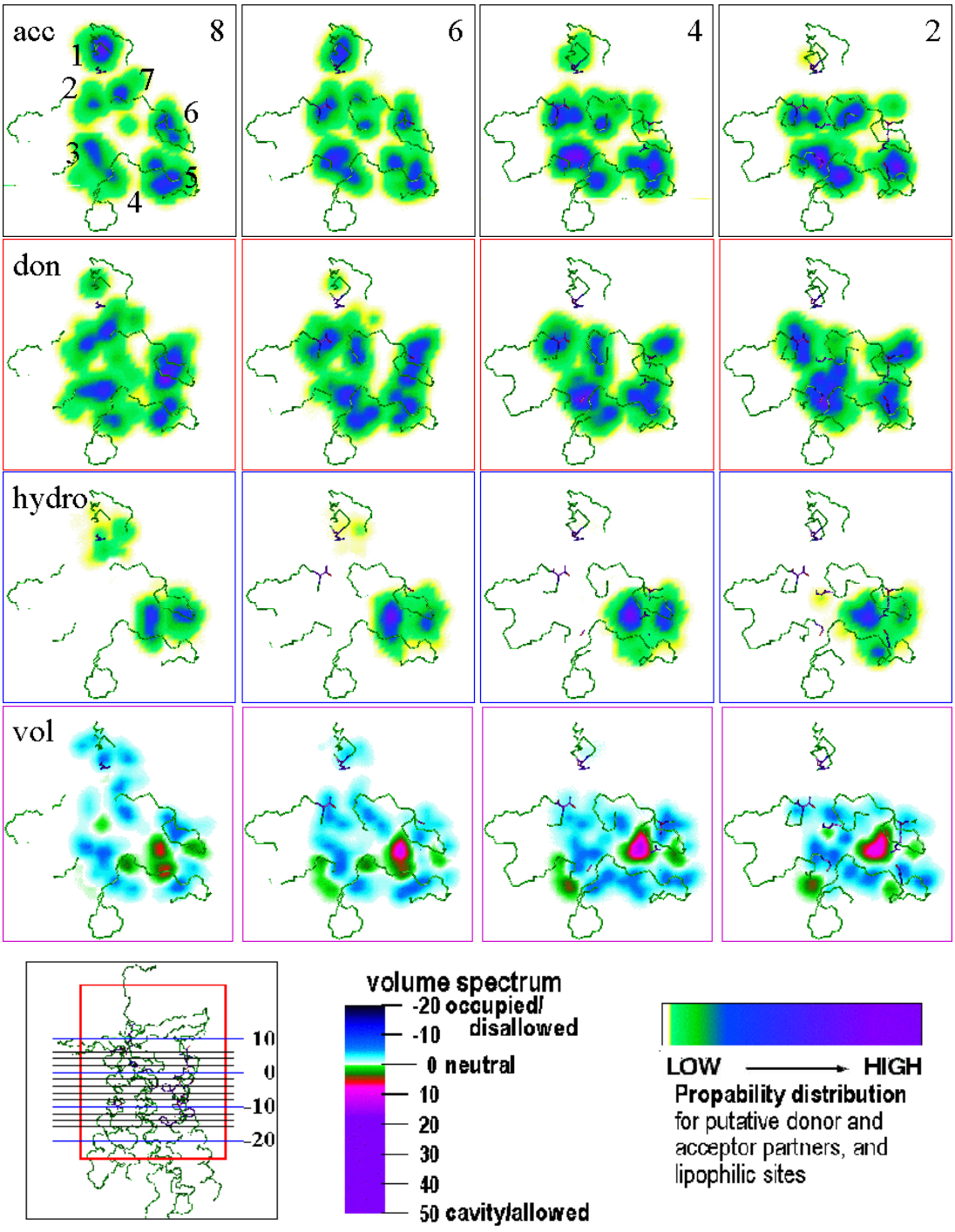


*Figure 6.3: Property distribution maps averaged over the 14 conformations of the human CCK-B receptor calculated by DRAGHOME for putative acceptor (*acc*) and donor partners (*don*), lipophilic sites of the accessible volume (*hydro*) and the allowed and disallowed volume (*vol*). The positions of the maps are depicted in the square at the bottom left. The map at +8 Å is in the extracellular part of the receptor, the map at 0 Å approximately in the center of the binding site and the map at -18 Å is in the intracellular part of the receptor. The colour legends are also depicted at the bottom.*

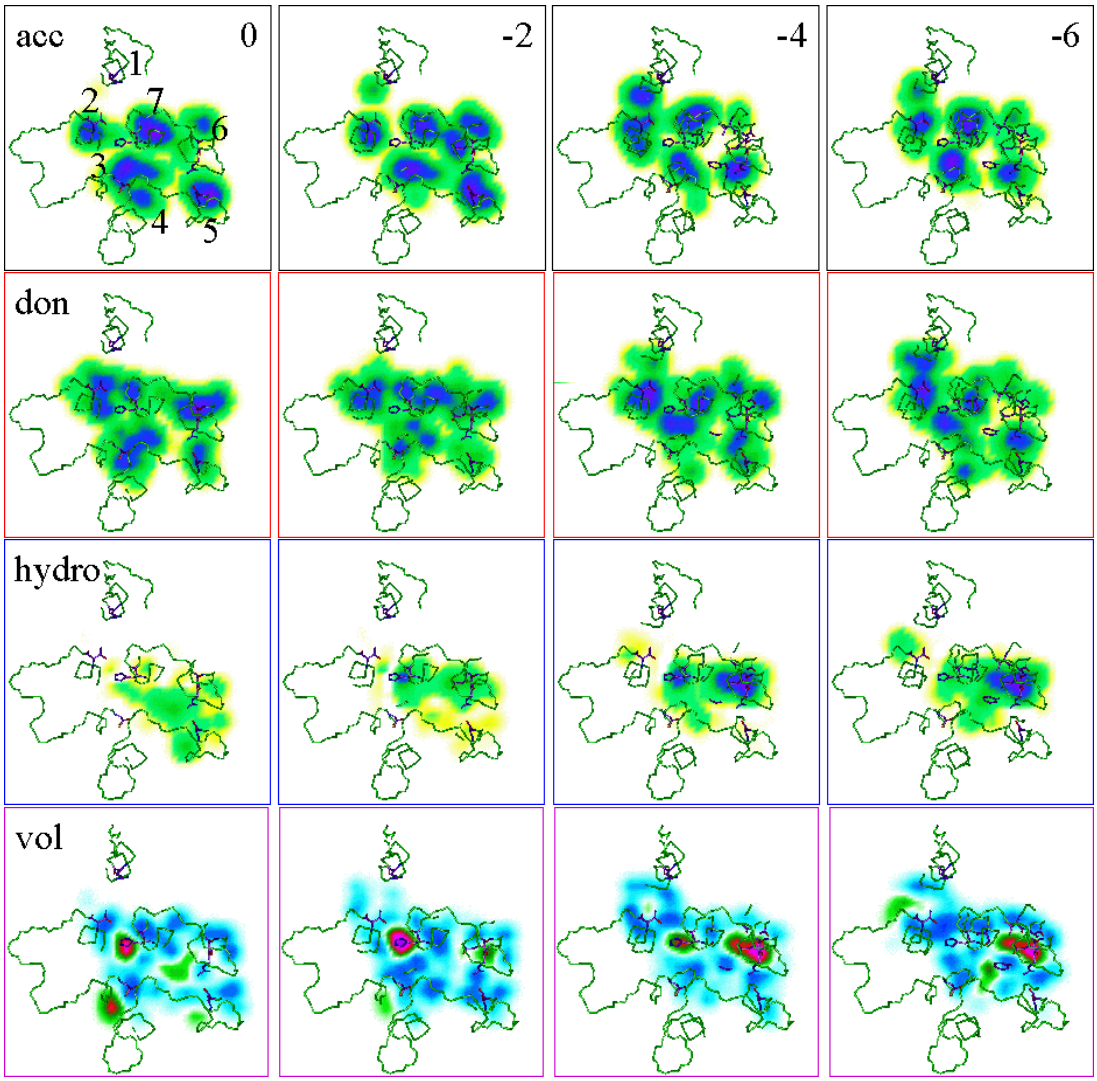


*Figure 6.4: Property distribution maps averaged over the 14 conformations of the human CCK-B receptor calculated by DRAGHOME for putative acceptor (*acc*) and donor partners (*don*), lipophilic sites of the accessible volume (*hydro*) and allowed and disallowed volume (*vol*). The positions of the maps and the colour legends are depicted in Figure 6.3.*

conformations of the receptor protein. However, these property distributions of the receptor are used for the subsequent alignment to match with complementary molecular properties of the ligands.

6.3.3.2 Docking with TORSEAL Alignment

In order to start with a reasonable conformation, the ligands were modelled with a conformation based on the crystal structures found for some representatives in the CSD (see also Figure 5.14, Chapter 5). All ligand molecules were then aligned to compound **89**. This compound shows the most rigid skeleton in the data set and it is assumed that all ligands bind with a similar overall shape to the same binding site. All benzodiazepine derivatives could be well aligned to compound **89** using TORSEAL after a preorientation of the benzodiazepine part with respect

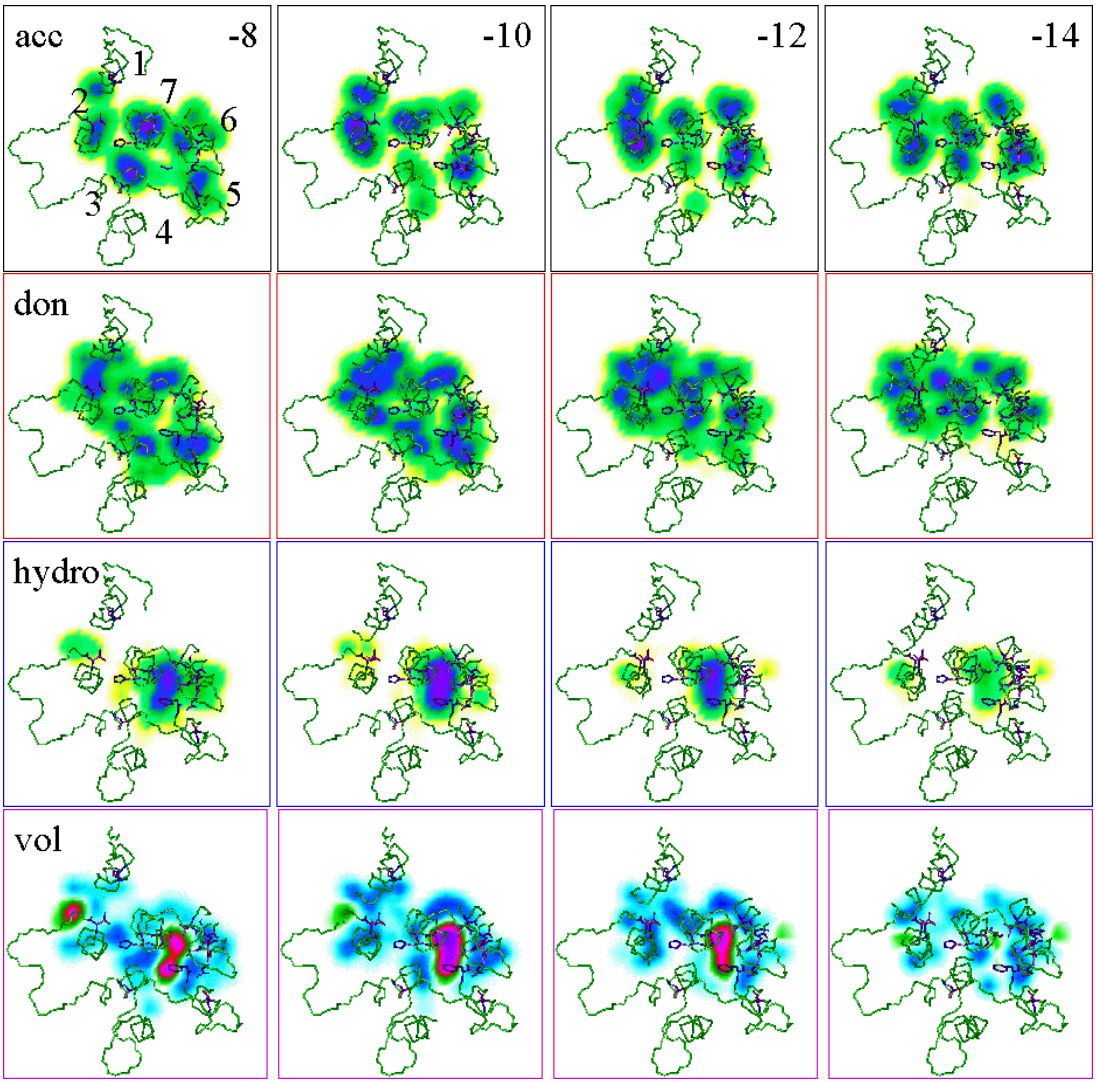


*Figure 6.5: Property distribution maps averaged over the 14 conformations of the human CCK-B receptor calculated by DRAGHOME for putative acceptor (*acc*) and donor partners (*don*), lipophilic sites of the accessible volume (*hydro*) and allowed and disallowed volume (*vol*). The positions of the maps and the colour legends are depicted in Figure 6.3.*

to the reference has been applied. Only the R-enantiomer of L-364,718 (**66**), was somewhat less well aligned ("black" structure in Figure 6.6A). This is due to the lack of an urea group within the compound. The S-enantiomers of the benzodiazepine derivatives could be aligned to compound **89** after a ring flip analog to the suggestion of Van der Bent et al. [208] described in Section 5.6 and Figure 5.13. However, the quinazolinone derivatives, e.g., compound **52**, (Figure 6.6) and the diphenylpyrazolidinone derivatives, e.g., compounds **57** and **58**, (Figure 6.6C) are difficult to align with compound **89**. It is not obvious which ring has to be aligned to which position of compound **89** and also TORSEAL suggests several solutions of comparable significance and similarity.

The best solutions selected by visual inspection of the first initial rigid docking trial with DRAGHOME are depicted in Figure 6.7. It can be seen that the cavities 2 and 3 (the cavity between the helices 3, 4, 5, 6, and 7 in maps -4 to -12 Å and the cavity between helices 1, 2,

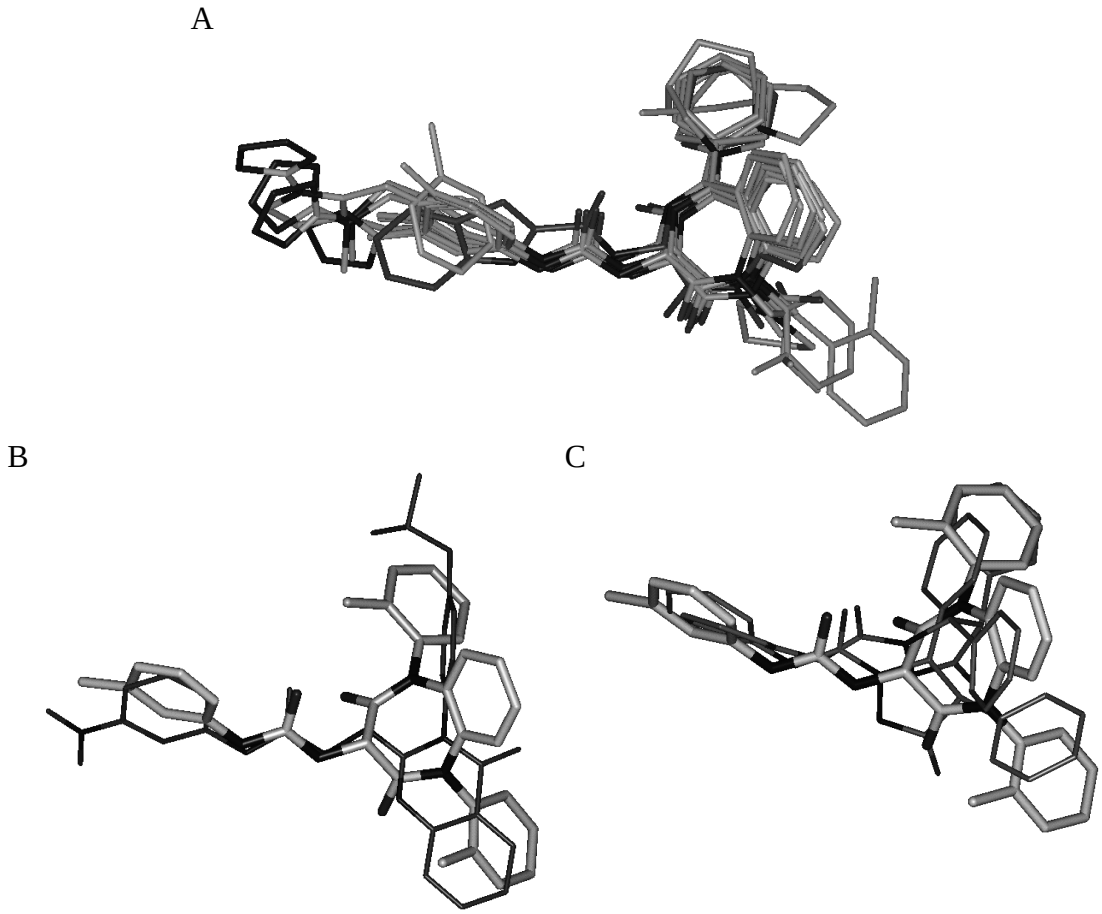


*Figure 6.6: TORSEAL alignments of several benzodiazepine derivatives with compound **66** in black (A) and of a selected quinazolinone derivative (compound **52**) with compound **89** (stick-mode structure) (B) and two enantiomers of a diphenylpyrazolidinone derivative (compounds **57** and **58**) with compound **89** (stick-mode structure) (C).*

3 and 7 in maps 0 to -4 Å, respectively, section 6.3.3.1) are almost completely occupied by the ligands. Only cavity one (the cavity between the helices 3, 4, 5, 6 and 7 in maps +8 to 0 Å) remained unoccupied. Furthermore, it is obvious that many important residues, e.g., His376, Phe342, Ser379, Phe227, Val349, Tyr350, Trp346 and Asn353 are located in close proximity to the two occupied cavities and are therefore in direct contact to the docked ligands.

Figure 6.8 depicts for five ligands the docking mode within the receptor with respect to compound **89**, whereby the membrane plane is oriented parallel to the paper plane. The benzodiazepine derivatives are fitting rather well on each other although two main positions can be determined. In the first position, the two substituents on the benzodiazepine ring are oriented in a plane parallel to the membrane plane as for compound **89** and in the second orientation those substituents are oriented in a plane orthogonal to the membrane plane as for compound **76** (Figure 6.8A). It is remarkable that compound **73** L-740,093 is the only benzodiazepine derivative which fails to fit with its urea group onto the corresponding group of compound **89** (Figure 6.8B). The quinazolinone and the diphenylpyrazolidinone derivatives are docked well in the receptor, but in none of the solutions the three substituents are fitted onto all the three substituents of compound **89**. Either the substituent attached to the urea-group together with

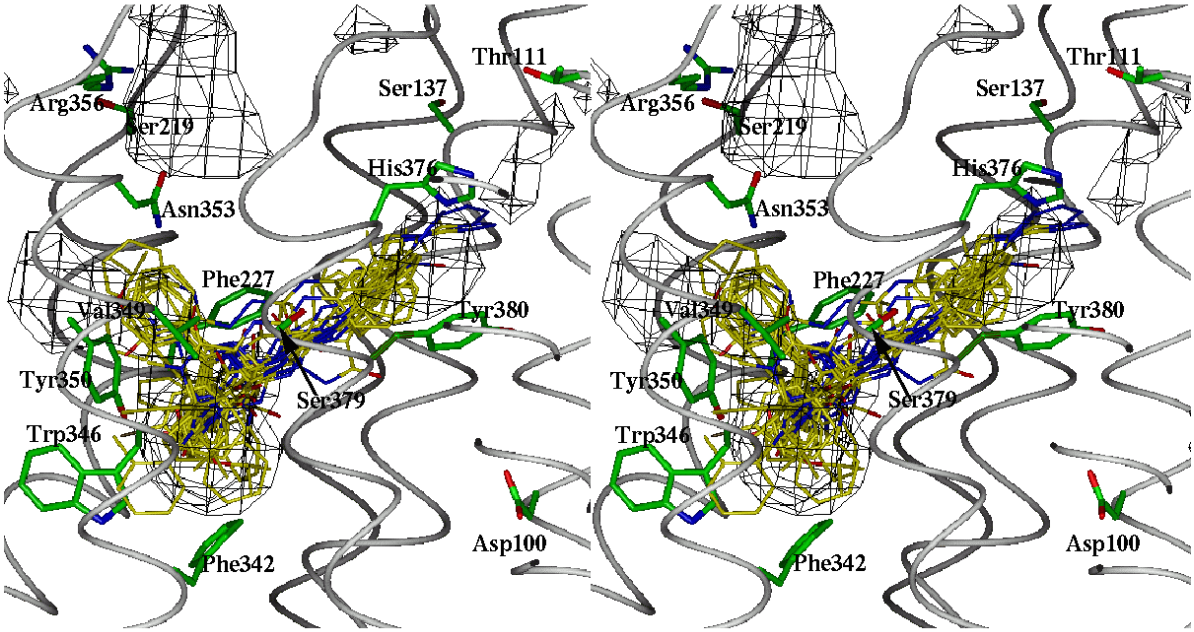


Figure 6.7: Side-by-side stereo presentation of the ligands docked in the first initial rigid docking trial with DRAGHOME in the binding site of the human CCK-B receptor model. A contour grid (at value 4 ∼ 90%) is depicted of the volume property of the protein, together with key amino acid residues, Tyr380, and the ribbon of the protein.

one of the other rings fits reasonably well, while the other ring does not fit at all (compound **57**, the light structure in Figure 6.8C, and compound **53**, Figure 6.8D), or the three aromatic rings are fitted approximately without matching the urea group (compound **58**, the dark structure in Figure 6.8C). These last results confirm the observations with the TORSEAL alignment, that the alignment of these classes of derivatives to compound **89** is difficult, since they do not share exactly the same structural properties with compound **89**.

6.3.3.3 Docking with MOCCA Conformations

Up to this stage of the investigation, the flexibility of the ligands were only considered in the prealignment step, not during the docking procedure. However, in order to assess whether better results can be obtained using alternative conformations of the ligands, especially when referring to the quinazolinone and the diphenylpyrazolidinone derivatives (see considerations of section 6.2.2), MOCCA calculations and analyses [235] were performed and subsequently docked with DRAGHOME. These analyses resulted in 900 conformations for each compound. Reduction of the number of conformations of each ligand and thereby decreasing the number of docking experiments could be achieved by some kind of conformational filtering, as, e.g., by selecting representative conformations with NMRCLUST [231]. However, this was not pursued in this study, in contrast all 900 conformations were used in the subsequent docking procedure. The minimum and maximum energy found by the MOCCA analyses for every ligand is listed in Table 6.7.

compound	min. energy [kcal/mol]	max. energy [kcal/mol]	# rule 1	best rank rule1	energy best rank	# rule 2	best rank rule 2	energy best rank	# rule 3	best rank rule 3	energy best rank	minim.	rank of conf.
31 L-364,718-flipped	149.536	154.584	0	1258									
66 L-364,718-R	148.805	152.557	61	4	150.249							1	4
66 L-364,718-R-flipped	152.829	157.945	0	1774									
71 L-365,260-S-flipped	187.544	193.938	0	1197									
71 L-365,260-S	142.269	147.809	57	1	144.214								
72 L-365,260	162.814	167.397	8	3	165.404							1	3
73 L-740,093	176.518	181.705	92	1	178.832							1	1
73b L-740,093-S	162.004	167.741	0	329	163.464								
74 L-708,474	156.986	163.266	17	15	159.007							1	15
75 YM-022	203.804	212.749	100	1	206.804							1	1
75 YM-022-flipped	208.390	217.613	1	58	209.691								
75b YM-022-S -flipped	214.089	223.490	13	8	216.123								
75b YM-022-S	184.886	193.826	7	58	187.253								
76 L-368,935	173.747	182.857	100	1	175.556							1	1
77 L-738,425	115.418	122.628	98	1	118.492							1	1
77 L-738,425-flipped	115.883	122.597	36	17	117.642								
78	129.801	136.099	98	1	131.976							1	1
79	113.472	120.366	88	1	116.962							1	1
80	120.752	127.064	94	3	123.818							1	3
81 L-736,380	112.633	119.147	98	1	113.645							1	1
82	148.591	153.843	18	1	151.028							1	1
83	166.288	174.703	96	1	167.410							1	1
57 LY 288512	120.702	126.898	0	199		44	1	123.708	1	70	121.839		
58 LY 288513	118.377	124.448	0	323		1	42	119.660	85	1	120.981		
60	124.790	130.614	0	1204		11	12	129.721	11	1	127.964	1	1
61	128.900	136.139	0	699		7	1	132.770	0	277	130.909		
44 LY 247348	127.161	136.539	2	54	130.014	5	55	131.210	17	3	130.551	1	3
48	130.117	143.736	2	26	143.737	3	1	134.190	16	8	133.163	1	8
49	131.762	140.559	0	147		1	71	136.979	3	33	138.385		
50	128.710	137.009	0	1069		0	188	132.987	18	4	132.127	1	4
51	105.328	114.523	0	210		2	38	109.768	10	23	109.491	1	23
52	193.597	201.666	0	310		5	1	198.759	18	10	194.897	1	10
53	198.685	207.244	0	461		0	216	204.312	73	8	200.699	1	8
54	191.602	199.201	0	320		0	570	196.251	83	1	194.313	1	1
55	128.592	135.736	0	1073		1	97	131.030	77	3	131.569	1	3
56	132.121	138.370	1	98	135.934	2	10	136.336	3	30	136.405	1	30
84	170.568	178.506	27	10	175.075							1	10
85	153.075	161.184	69	4	155.341							1	4
87	145.672	151.310	93	2	147.791							1	2
88	140.719	147.045	93	1	143.911							1	1
89	247.648	254.887	86	1	250.366							1	1
89-flipped	230.373	237.620	0	1377									
90	101.983	111.231	99	1	104.375							1	1
92	153.864	162.730	97	1	159.439							1	1
97	153.683	165.712	87	1	158.849							1	1
98	97.129	106.483	100	1	99.042							1	1
100	147.136	153.910	100	1	150.751							1	1
100-flipped	140.083	147.646	11	3	142.402								
CCK-4	69.715	83.540											
CCK-8	134.333	161.100											

Table 6.7: Minimum and maximum energies of the MOCCA analyses and the results of the docking selection, see text for details.

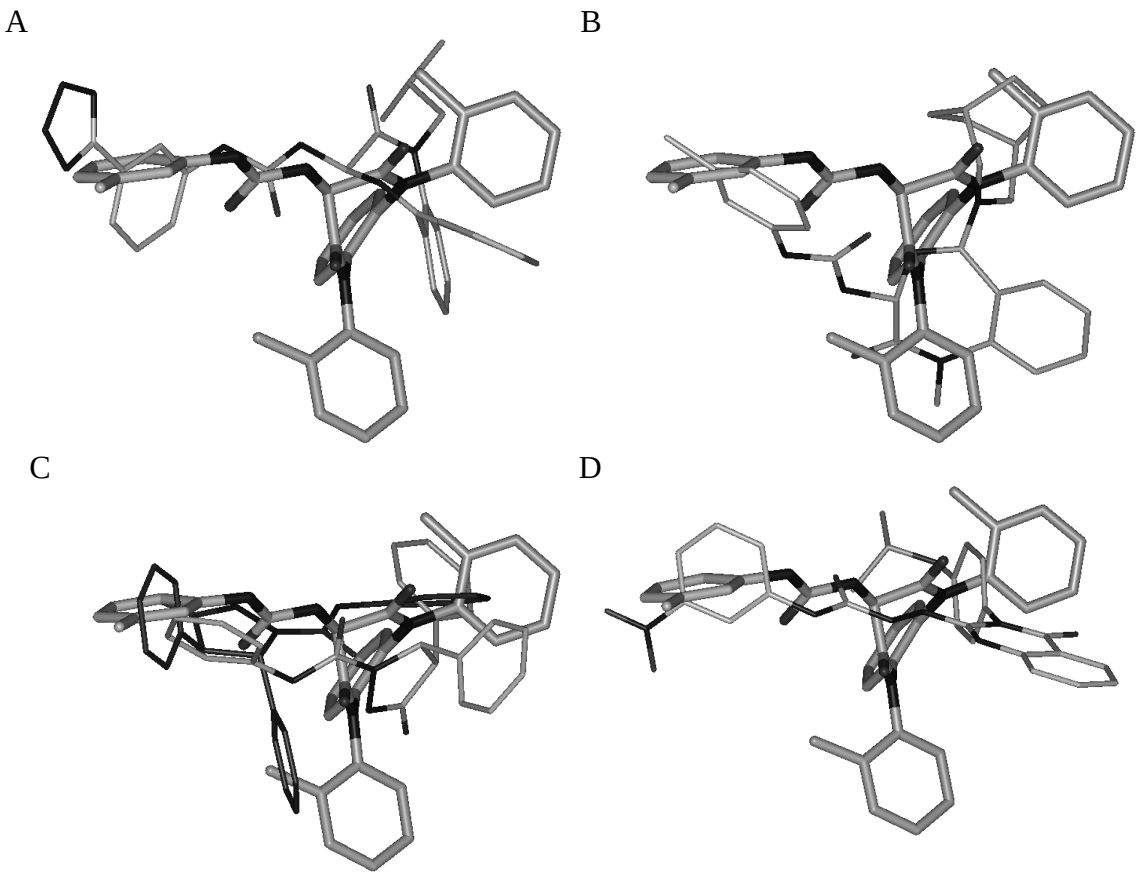


*Figure 6.8: Several ligands docked by DRAGHOME are depicted with respect to compound **89**. (A) compound **76**, L-368,935; (B) compound **73**, L-740,093; (C) compounds **57** (*dark*) and **58**, LY 288512 and LY 288513, respectively; (D) compound **53**.*

6.3.3.4 Analyses of Docking Results

Selection Method One

The top 10 solutions of several ligands are depicted in Figures 6.9 to 6.11 after docking. These top 10 solutions were selected by applying the strategy of the average of the weighted rankings of the four properties.

It can be seen (Figures 6.9) that all ten solutions for compound **76** L-368,935 have similar positions, whereby the urea group is located between helices 3 and 7 in proximity to Ser379. This is a similar position as found in the first rigid docking trial with DRAGHOME. It can be hypothesised that for a stable complex, the urea group, which is essential for CCK-B antagonists (see Chapter 5), should adopt a position between helices 3 and 7 in close proximity to Ser379. Therefore, the so-called "urea-group"-rule (rule 1) is applied to select the solutions out of the top 100, which have their urea oxygen atom within a distance of 3 Å of the position of the urea oxygen of compound **76** L-368,935. These countings for all the compounds are listed in Table 6.7, together with the rank and the conformational energy of the compound with the highest rank obeying the selection-rule.

The top 10 of the R-enantiomer of L-364,718, compound **66**, has four solutions comparable to compound **76**, but it displays also solutions which are considerably different (Figure 6.9). The

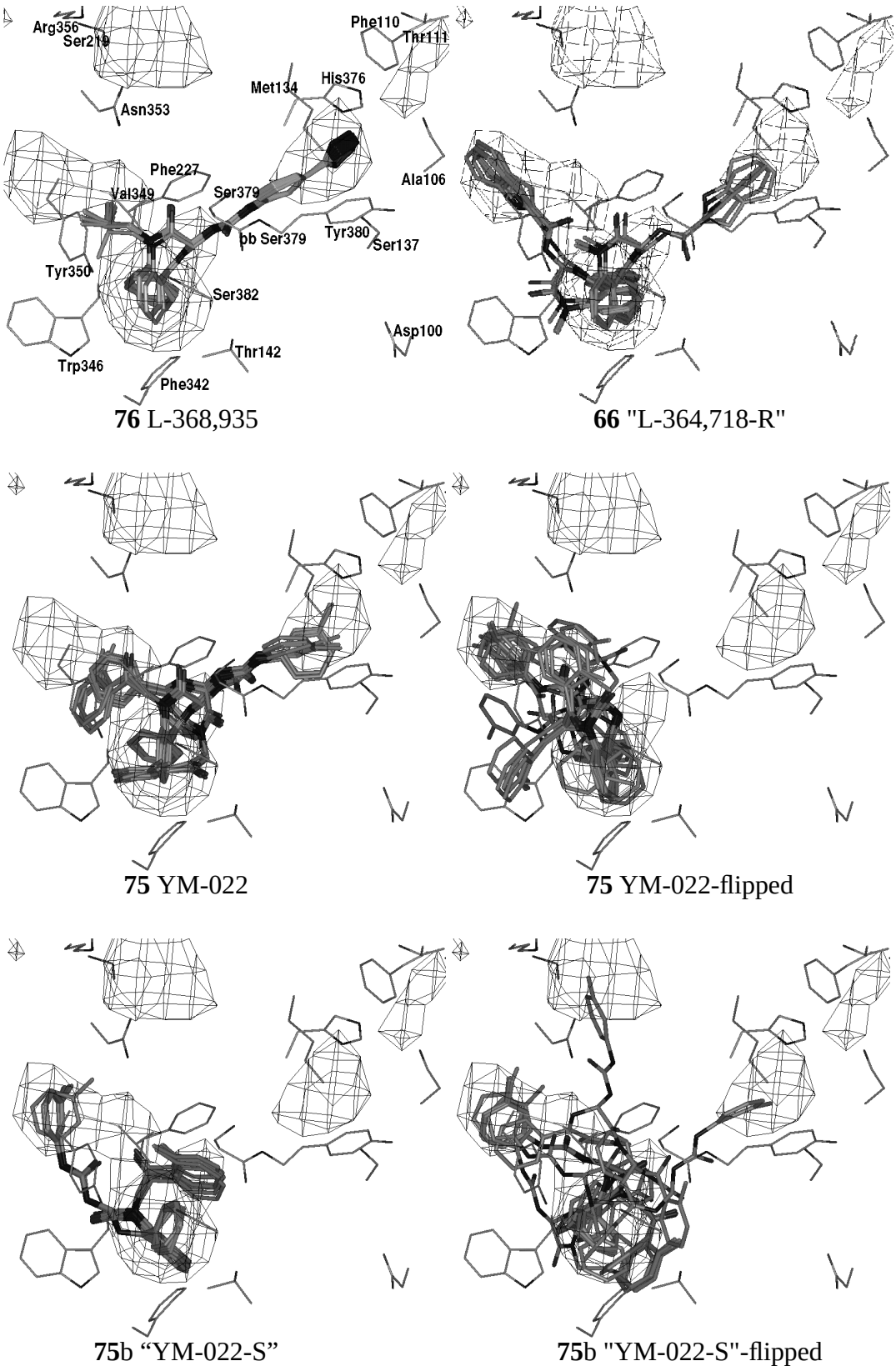


Figure 6.9: The top 10 solutions of several ligands docked with DRAGHOME.

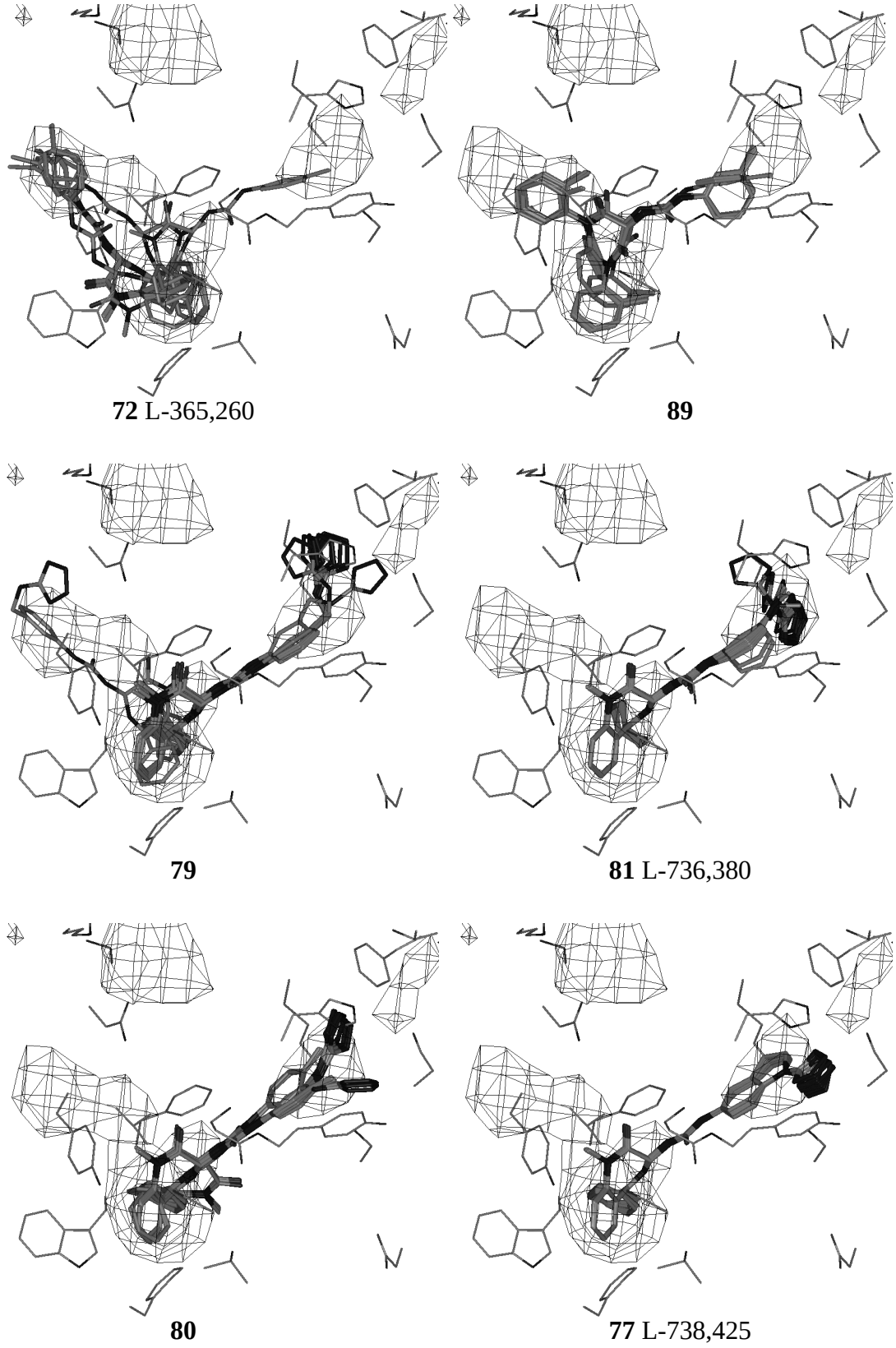


Figure 6.10: The top 10 solutions of several ligands docked with DRAGHOME.

latter ones occupy only the second cavity (the cavity between the helices 3, 4, 5, 6, and 7 in maps -4 to -12 Å, section 6.3.3.1). The urea substituents of these structures are located between helices 5 and 6 (the most left side of the second cavity in Figure 6.9) and they are almost in contact with the surrounding membrane by penetrating the helix 5-helix 6 interface. Similar orientations were also solutions in the first rigid docking trial with DRAGHOME, however on basis of the mutual spatial similarity strategy of Schafferhans [221] they were then discarted as good solutions.

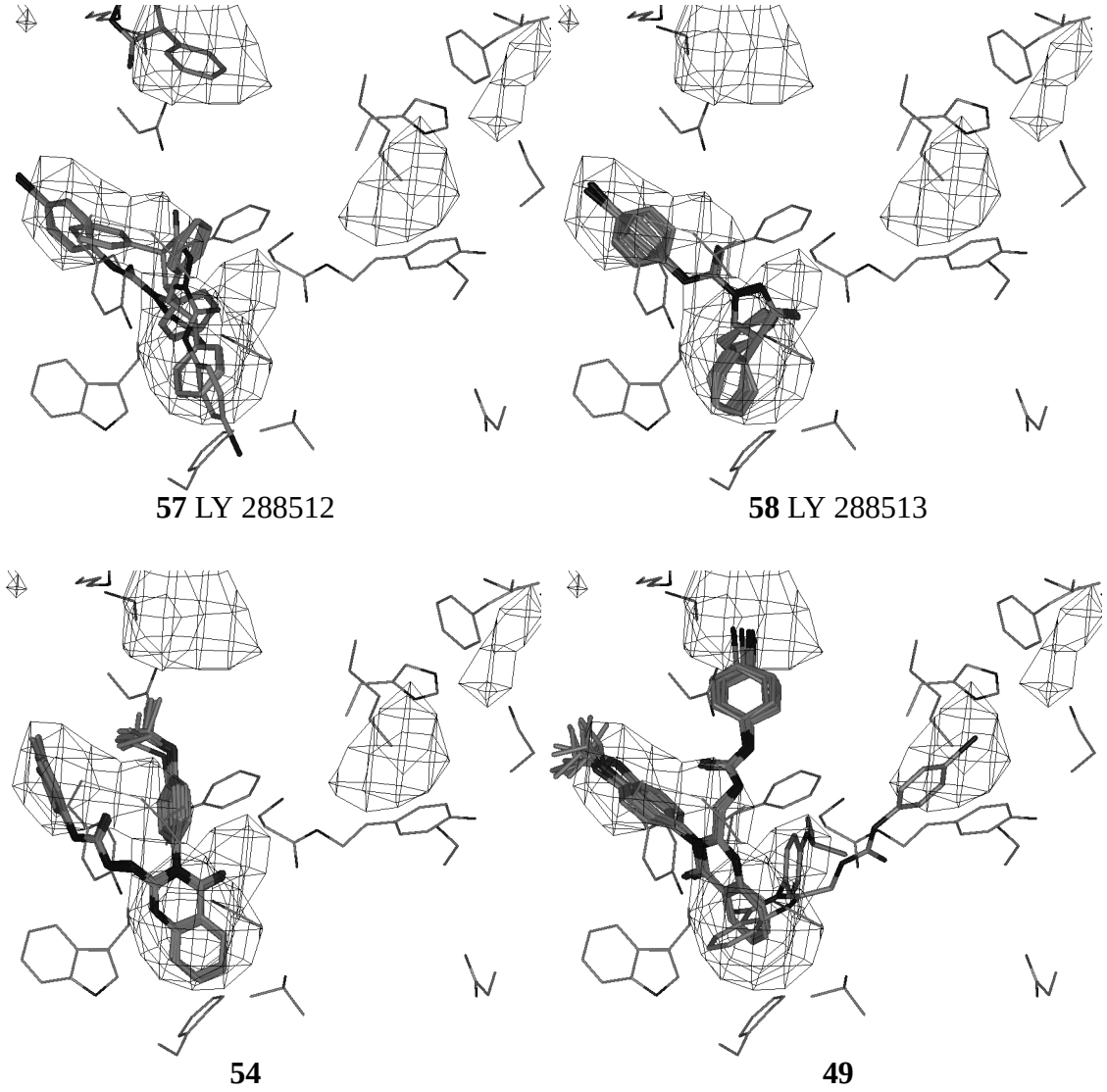


Figure 6.11: The top 10 solutions of several ligands docked with DRAGHOME.

It was found that all the "flipped" conformations of the benzodiazepine derivatives could not be docked satisfactory in the binding site of the receptor. Not a single of these conformers adopted a position that showed some similarity with that of compound **76**. It is remarkable that the top 10 solutions obtained for these "flipped"-conformations exhibited a significant divergence. In Figure 6.9, the top 10 solutions for the different conformations and enantiomers of YM-022 are given as an example for the influence of stereoisomerism and the "flipped"-conformation. Although for the R-enantiomer in the "non-flipped"-conformation, two main orientations are

found, with mirrored alignments of the urea-group. It can be seen that the solutions occupy the cavities two and three and that the urea group of these solutions are in close proximity to Ser379. The S-enantiomer in the "non-flipped"-conformation adopts only one position, similar to that of compound **66**. The "flipped"-conformations of both enantiomers are scrambled through cavity two, exceeding beyond the border of this cavity.

It is remarkable that only one solution of L-365,260, compound **72**, the prototype CCK-B antagonist, has a somewhat similar position to that of compound **76** L-368,935 (Figure 6.10). Compound **89**, the template structure of the first rigid docking trial with DRAGHOME shows solutions between helices 3 and 7, comparable to the solutions of compound **76** L-368,935. However, the two ring-substituents are oriented in two different positions (Figure 6.10).

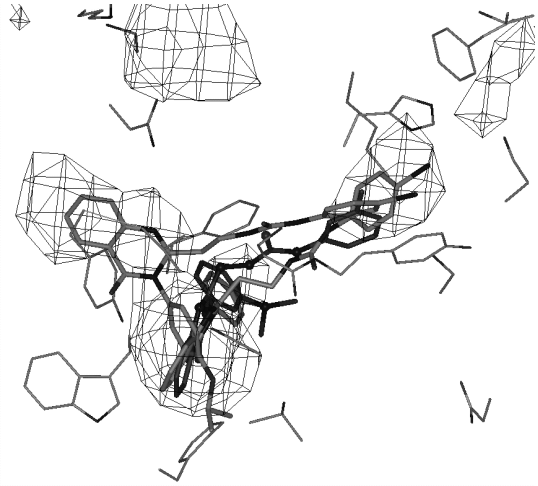
There are several ligands with a tetrazol-5-yl amino group attached to the urea substituent. Although this tetrazol-5-yl amino group adopts slightly altered positions (Figure 6.10), it can be seen that it almost always occupies cavity three (**77**, **79**, **80**, and **81**). This tetrazol-5-yl amino group is "sandwiched" by the two aromatic residues His376 and Tyr380, and surrounded by the residues Phe110, Ala106 (oxygen-backbone-atom), Met134 and Ser137.

It can be seen from Table 6.7 that several of the benzodiazepine derivatives have high counts for the urea-group-selection rule. It is remarkable that these in general are also the structures possessing good binding affinity. In contrast to this result, their "flipped"-conformations show low counts and also the S-enantiomers of YM-022 (**75**), L-740,093 (**73**) and L-364,718 (**31**) have low counts. L-364,718 (**31**) was experimentally determined to exhibit a low binding affinity for the CCK-B receptor. For the other compounds only data for the R-enantiomers were reported. Only the prototype CCK-B antagonist, compound **72** L-365,260, has better countings for its S-enantiomer which has been reported to bind with lower affinity than the R-enantiomer to the CCK-B receptor.

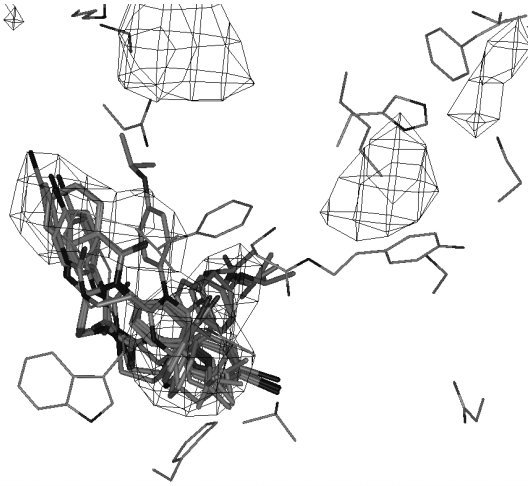
The quinazolinone derivatives (e.g., compounds **57** LY 288512 and **58** LY 288513, Figure 6.11) and the diphenylpyrazolidinone derivatives (e.g., compounds **54** and **49**, in Figure 6.11) have only few solutions in which their structure occupies both cavities two and three. It is remarkable that for compound **58** LY 288513, the enantiomer with the better binding affinity, only one position is found and that in contrast for compound **57** LY 288512, the enantiomer with the lower binding affinity, several different positions were identified. When all the top 10 solutions of all the quinazolinone and diphenylpyrazolidinone derivatives were analysed, it was observed that many solutions have an oxygen atom in proximity to Ser379, however not in an overlapping mode to the urea group of compound **76** L-368,935. Therefore the above-mentioned "urea-group"-selection rule was slightly extended to yield two further selection criteria, rule 2 and 3, respectively. In the second selection rule, the distance between the position close to Ser379 and the urea oxygen (or sulfur atom) of the compound was measured and the number of solutions out of the top 100 with a distance smaller then 3 Å were counted. In the third selection rule the oxygen atom of the ring system (quinazolinone or pyrazolidinone) was selected. These countings are also listed in Table 6.7. It is obvious from Table 6.7, that only a few solutions obey the first selection rule, but that the countings for the two extra rules are higher.

In Figure 6.12, several solutions of compound **48** and compound **56** obeying the first selection rule are shown. Their urea groups are located between helices 3 and 7, similar to the urea group position of compound **76** L-368,935.

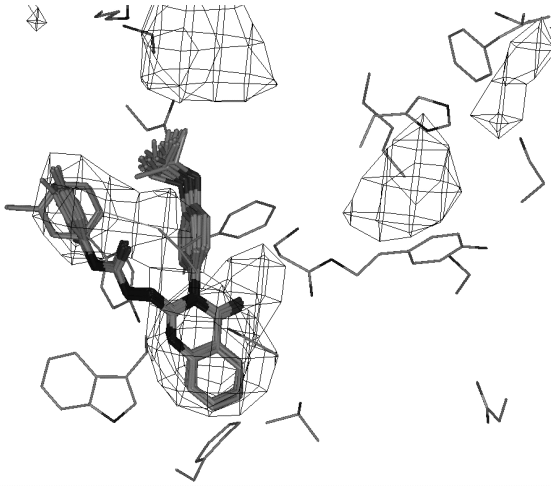
When the solutions of the different quinazolinone compounds, obeying the third selection rule, are analysed, it can be seen that many solutions are located in similar positions (Figure 6.12, **48**, **54** and **56**). The solutions obeying the second selection-rule are not only lower in countings,



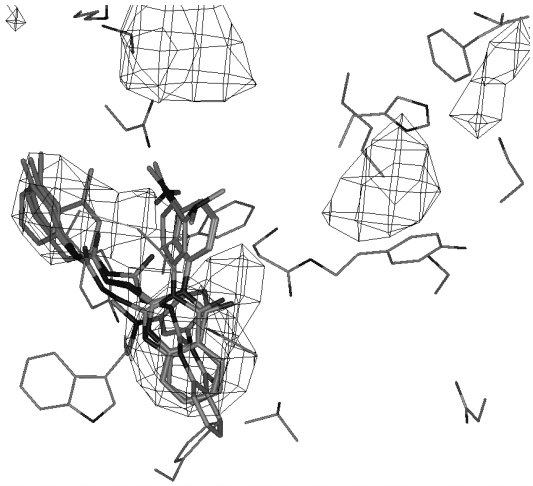
56 (*ball and stick*) and **48**, 1^{st}-selection rule



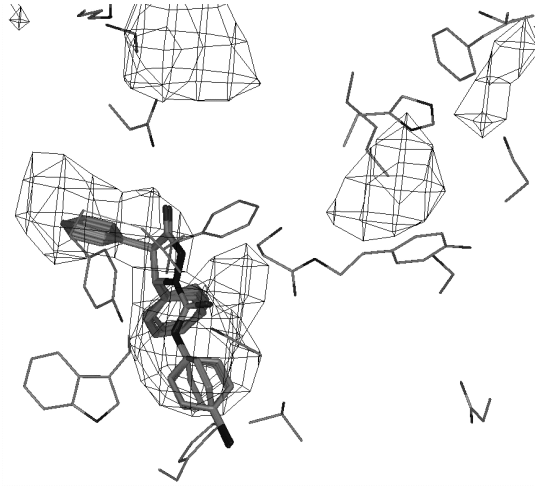
48, 3^{rd}-selection rule



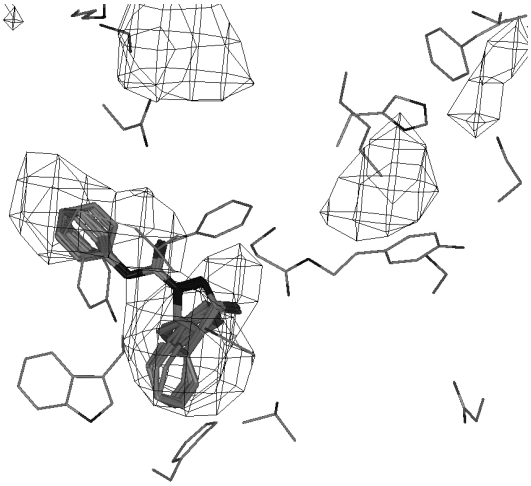
54 , 3^{rd}-selection rule



56 , 3^{rd}-selection rule



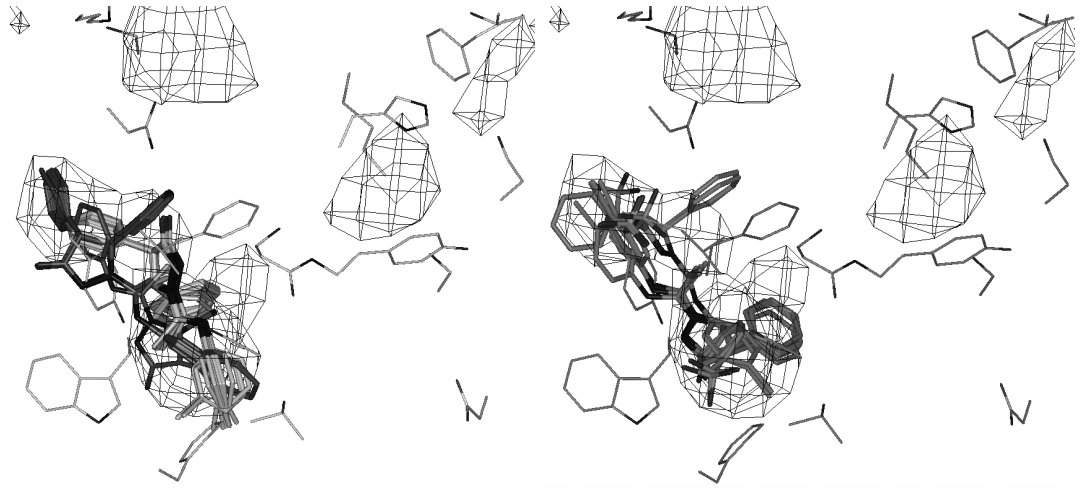
57 LY 288512 , 2^{nd}-selection rule



58 LY 288513, 3^{rd}-selection rule

Figure 6.12: The solutions of several quinazolinone and diphenylpyrazolidinone derivatives obeying the different selection rules.

they also do not have always very similar positions. Even when the solutions were very different they almost never occupied the first or third cavity.
The diphenylpyrazolidinone derivatives, compounds **57** LY 288512 and **58** LY 288513, also have more countings with respect to the second and third selection rules. However, it is remarkable that compound **57** LY 288512, the weaker binding enantiomer, obeys the second selection rule and has its urea-oxygen atom in close proximity to Ser379. In contrast, compound **58** LY 288513, the strong binding enantiomer has its ring-oxygen atom in close proximity to Ser379 and thus follows the third selection rule.



60, 2^{nd} (*dark*) & 3^{rd}-selection rule **61**, 2^{nd}-selection rule

*Figure 6.13: Solutions of the diphenylpyrazolidinone derivative compounds **60** and **61** obeying the different selection rules.*

Compound **60** is a diphenylpyrazolidinone derivative with RS-enantiomery (see Figure 5.11). This weak binding compound has solutions obeying both rules, but none is really well docked, since atoms exceed beyond cavity 2 (Figure 6.13). Compound **61**, the stronger binding SR-enantiomer, obeys only the second rule, however the solutions are better docked in comparison to compound **60**. Therefore, a better binding of the SR-enantiomer can be hypothesised.
The results described here suggest that the binding mode for the quinazolinone and the diphenylpyrazolidinone derivatives is different to that compared with the benzodiazepine derivatives. However, their binding sites are greatly overlapping in the second cavity of the receptor.

Selection Method Two

The second selection of the best solutions created by DRAGHOME is done using the method of grouping the solutions based on the RMSD of the non-hydrogen atoms of the ligand.
In the first grouping of the solutions of compound **76** L-368,935 a threshold RMSD value of 7 Å was used and 40 groups were obtained. The first two groups have both the highest number of members in the group, the highest number of hits (DRAGHOME can find a solution more than once in a docking trial and this amount for the total group is the number of hits) and the lowest average of the total scores of the members. The first member of these two groups are depicted in Figure 6.14A. The first member of the second group is in an orientation that is similar to the orientations also found for compound **76** L-368,935 in the initial rigid docking trial and in

the first selection method. In this orientation the ligand occupies cavities 2 and 3 (the cavity between the helices 3, 4, 5, 6, and 7 in maps -4 to -12 Å, and the cavity between helices 1, 2, 3 and 7 in maps 0 to -4 Å, respectively, section 6.3.3.1). In contrast, the first member of the first group occupies only cavity 2 and one of its substituents is positioned between the helices 5 and 6, an orientation also found in the first selection method for other compounds.

In Figure 6.15A it can be seen that the spread in total score of the different members of a group can be large. Thus, solutions which are positioned in the same area within the binding site of the receptor have rather different total scores. In contrast, the different groups can have similar distributions in the total scores of their members. However, the first member of all groups have clearly different orientations, since the RMSD values in the distance matrix of the first member of all groups are in general larger than 7 Å (Figure 6.15B). The few RMSD values below 7 Å are obtained through rounding off and are close to 7 Å.

The spread in the total score of the different members of a group is in the second grouping smaller than in the first grouping, but is still significant (Figure 6.15C). That the 55 groups of the second grouping have different orientations can be seen in the distance matrix of Figure 6.15D. The distance matrix shows that the RMSD value of the first member of each group to the first member of each other group is larger than 2 Å (only a few yellow boxes caused by rounding up).

The distance matrix of group 2 in Figure 6.15E shows that the RMSD values within one group are in general not larger than 2 Å. Only a few RMSD values are larger than 2 Å (0.1% of all RMSD values and the largest RMSD value is 2.36 Å). However, the members of group 2 do not have identical total scores. The total score of the members of group 2 varies between -23363.0 and -28103.1. Thus, small deviations within the orientation of the ligand are responsible for considerable differences in the total score.

The best three groups of the second grouping (groups 2, 3 and 5) have the highest number of members, are in the top 4 of the highest number of hits, and are in the top 10 of the lowest averages of the total scores. The first member of these three groups are depicted in Figure 6.14B. These three stuctures have orientations occupying cavities 2 and 3, and have their urea-group positioned in close proximity to Ser379. These orientations are similar to the orientations of the best solutions of compound **76** L-368,935 selected by the first selection method and also similar to the orientation of compound **76** L-368,935 found in the first rigid docking trial with DRAGHOME. In contrast, the three groups with the lowest averages of the total scores (groups 1, 31 and 50) are only occupying cavity 2 and have their urea-group not in close proximity to Ser379 (Figure 6.14C). Furthermore, the urea-group substituent of these three structures is positioned between helices 5 and 6, disrupting the interhelical interactions between these two helices. Groups 31 and 50 have only a few members and they have also only a small number of hits. Therefore, these orientations are probably not favourable orientations for compound **76** L-368,935. However, group 1 is the fourth group in the number of members and the second in the number of hits. This group can be seen as the fourth best group. However, it has a significant smaller number of members in comparison to the best three groups.

The three groups with their first member depicted in Figure 6.14B, have rather similar number of members. However, their number of hits and their averages of the total scores of their members are different. Group 5 with as first member the magenta coloured structure, has a significant larger number of hits than groups 2 and 3, with as first member the purple and the green coloured structures, respectively. However, the average of the total scores of group 5 is significant higher (-23529.8) than the averages of the total scores of groups 2 and 3 (-26494.1 and -24410.3,

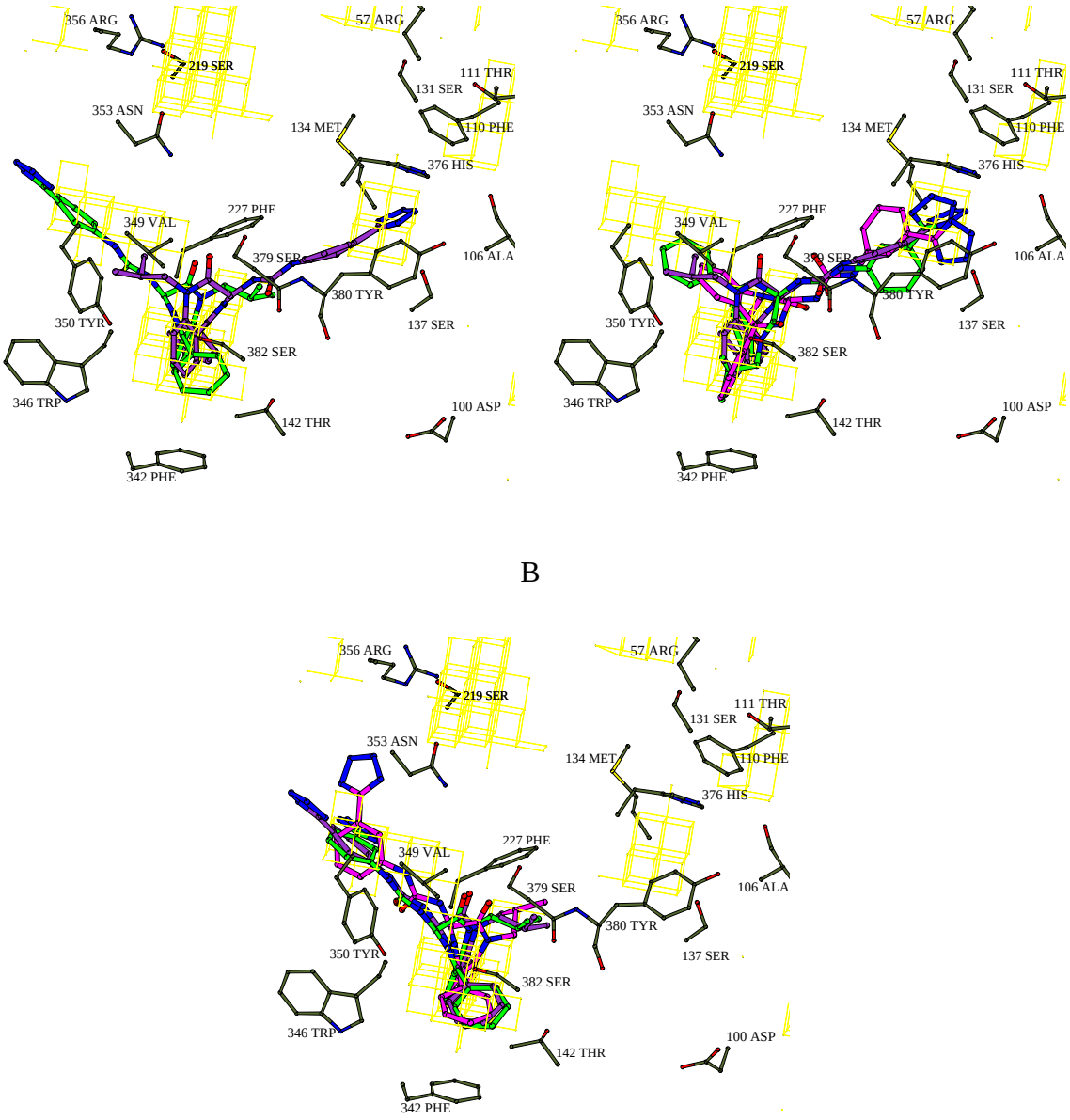


*Figure 6.14: (A) The first member of the best two groups of the first grouping of compound **76** L-368,935 with a RMSD threshold value of 7 Å (first member of group one = green compound, first member of group two = purple compound). (B) The first member of the best three groups of the second grouping of compound **76** L-368,935 with a RMSD threshold value of 2 Å. (C) The first member of the three groups of the second grouping of compound **76** L-368,935 with the lowest averages of the total scores of their members. (The pictures are made with the program MolScript [241].)*

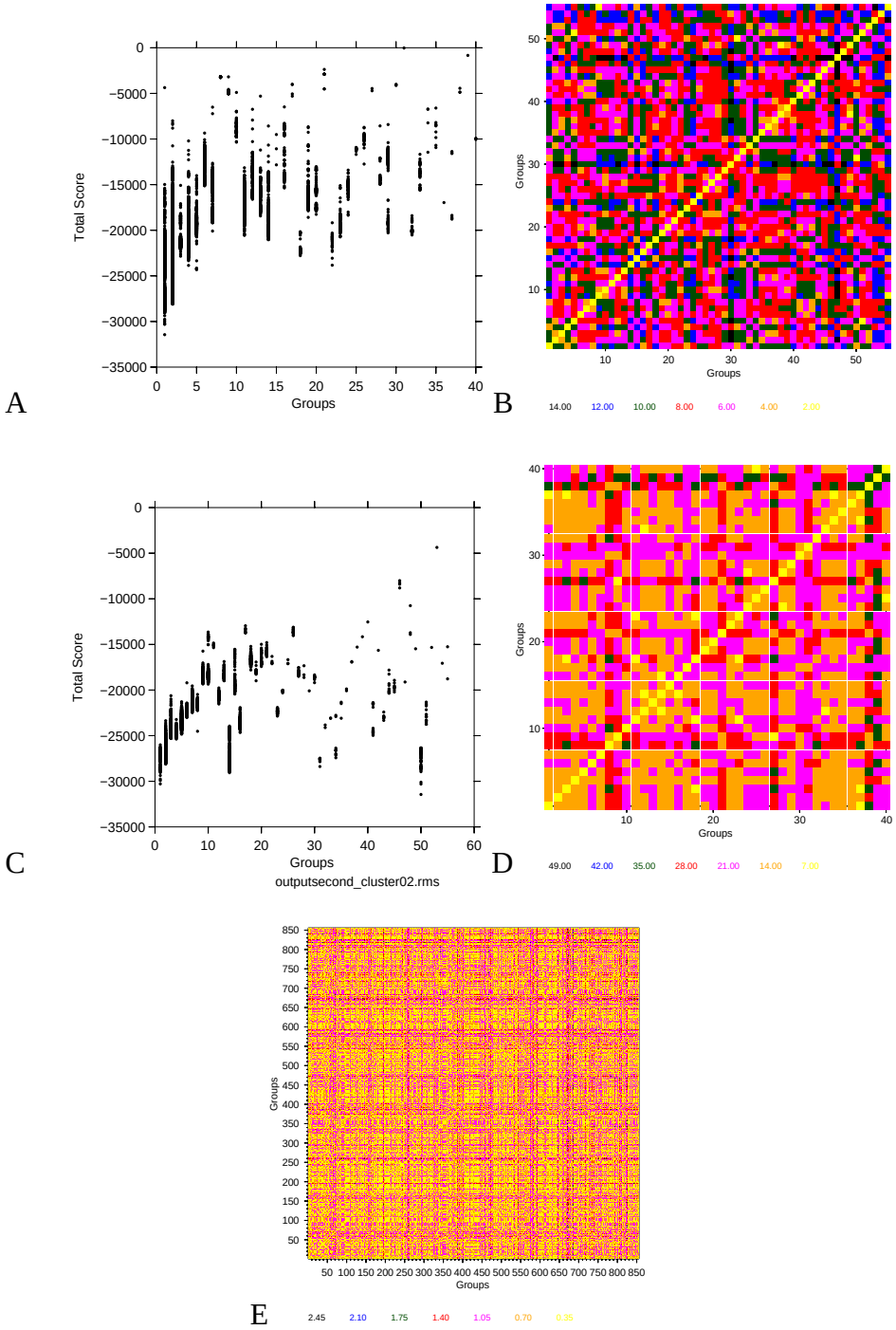


*Figure 6.15: Results of the two groupings of the solutions of compound **76** L-368,935. (A) The total scores of each solution versus the group number of which they are member after the first grouping with a threshold RMSD value of 7 Å. (B) The RMSD distance matrix of the first member of all groups after the first grouping. (C) The total scores of selected solutions from the first grouping versus the group number of which they are member after the second grouping with a threshold RMSD value of 2 Å. (D) The RMSD distance matrix of the first member of all groups after the second grouping. (E) The RMSD distance matrix of the 856 members of the second group after the second grouping.*

respectively). Group 2 was chosen as the best group of the three groups, since it has significantly the lowest average of the total score, and has, although not the highest number of hits, still a large number of hits (more than 9 times the number of members).

Similar thorough analyses can also be done for the other compounds.

For compound **75** YM-022 the two different enantiomers were both docked with conformations with their 3-amide side chain in the pseudo-axial position and in the pseudo-equatorial position (Figure 5.14, Chapter 5 [208]). It is remarkable that the grouping of the pseudo-axial conformations, the so-called "flipped"-conformations, of both enantiomers and the grouping of the pseudo-equatorial conformation of the S-enantiomer gave very large number of groups. It can be suggested that these three structures do not have a high preferred specific orientation within the binding site of the receptor and finding an orientation for them within the binding site of the receptor is difficult. The best groups of these three structures do not occupy cavity 3 and do not have orientations similar to the best group of compound **76** L-368,935. In contrast, the grouping of the conformations of the R-enantiomer with its 3-amide side chain in the pseudo-equatorial position gives a significant smaller number of groups and the best group is in a similar orientation as the best group of compound **76** L-368,935.

The compounds **57** LY 288512 and **58** LY 288513 have also a different number of groups. A larger number of groups was created for the enantiomer with the lower receptor affinity; compound **57** LY 288512. This same correlation between number of created groups and receptor affinity was also observed for the compounds **60** LY 294920 and **61** LY 294919, for the compounds **31** and **66** L-364,718 and for the compounds **71** and **72** L-365,260. This would suggest that structures with a smaller affinity have not a specific orientation within the binding site of the receptor or that it is difficult to find an orientation within the binding site for structures with a smaller binding affinity. Furthermore, compound **72** L-365,260 has its best group in an orientation similar to the best group of compound **76** L-368,935 and it has also a higher affinity than compound **71**, which has no group in an orientation similar to the best group of compound **76** L-368,935. However, for compound **66**, the enantiomer with the lower affinity, an orientation similar to the orientation of the best group of compound **76** L-368,935 was found for its third best group, whereas for compound **66** L-364,718, the enantiomer with the higher affinity, no group was found orientated in that way. This was also observed for compounds **57** LY 288512 and **58** LY 288513. Therefore, it remains difficult to draw correlations between the differences in binding affinity of the different enantiomers of the compounds and the observed binding modes.

The first member of the best group of the compounds also mentioned in the discussion of the first selection method, are depicted in Figure 6.16A. These are the compounds: **48**, **49**, **50**, **54**, **56**, **57** LY 288512, **58** LY 288513, **60** LY 294920, **61** LY 294919, **31** L-364,718, **66**, **71**, **71**-"flipped", **72** L-365,260, **75** YM-022 (both enantiomers and both enantiomers also "flipped"), **76** L-368,935, **77** L-738,425, **77**-"flipped" L-738,425-"flipped", **79**, **80**, **81** L-736,380, **89**, **90**, **97**, **100**, **100**-"flipped". In Figure 6.16B the first member of the best group are depicted for the same compounds, but now the best group is selected also on basis of the spatial similarity of various solutions generated for the different ligands for internal similarity consistency. This extra selection criterion was applied in analogy to the procedure applied by Schafferhans et al. [221]. Adding this selection criterion to the selection strategy means, that a group is selected as the best group, which has not only a high number of members, a high number of hits and a low average of the total scores of its members, but which also has an orientation similar to the

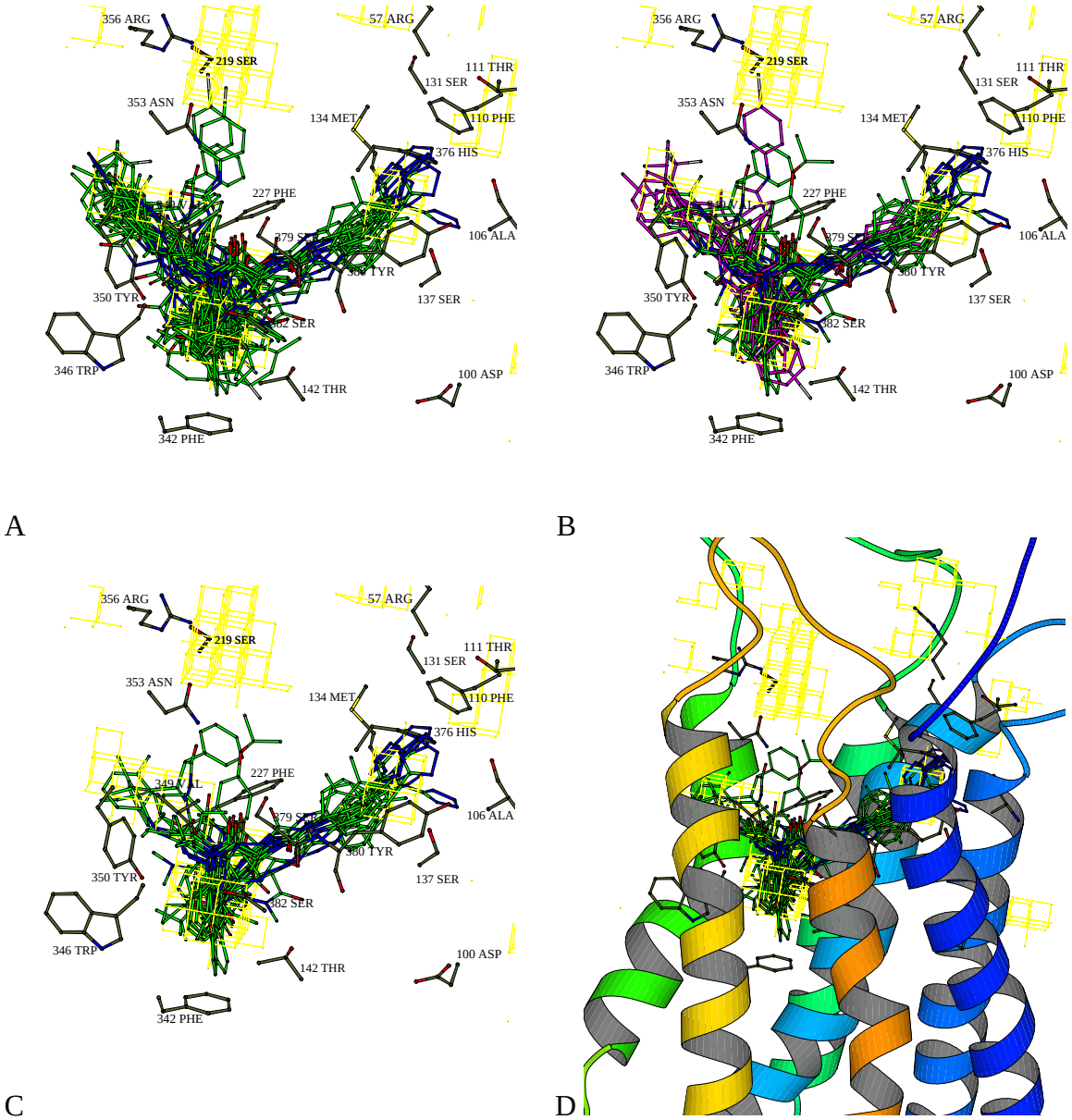


*Figure 6.16: (A) The first member of the best group of the compounds: **48**-**50**, **54**, **56**, **57** LY 288512, **58** LY 288513, **60** LY 294920, **61** LY 294919, **31** L-364,718, **66**, **71** (also "flipped"), **72** L-365,260, **75** YM-022 (two enantiomers and both also "flipped"), **76** L-368,935, **77** L-738,425 (also "flipped"), **79**, **80**, **81** L-736,380, **89**, **90**, **97**, **100** (also "flipped"). (B) The extra criterion of spatial similarity [221] was taken into account. However, compounds **71**-"flipped", **75** YM-022 (the two S-enantiomer conformations and the R-enantiomer with its side chain in the pseudo-axial position), **77**-"flipped" L-738,425, and **100**-"flipped" were left out. The compounds **48**, **49**, **57** LY 288512, **60** LY 294920, **61** LY 294919, **31** L-364,718, **66** and **71** have a binding affinity >100nM and are coloured magenta. (C) The magenta coloured structures are left out. (D) The compounds of (C) are positioned between the transmembrane helices in the upper part of the receptor. (The pictures are made with the program MolScript [241].)*

orientation of the best group of compound **76** L-368,935 with its urea-group in close proximity to Ser379 and occupying both cavities 2 and 3.
The so-called “flipped”-conformations (the 3-amide side chain in the pseudo-axial position) of the different compounds were left out in Figures 6.16B-C. These “flipped”-conformations gave orientations that were not following the spatial similarity criterion, or when they followed it, were less similar to the orientation of the best group of compound **76** L-368,935 than the orientations of the stretched conformations (the 3-amide side chain in the pseudo-axial position).
In Figure 6.16B it can be seen that by applying this spatial similarity criterion, the lefthand side of cavity 2, the area between helices 5 and 6, is occupied in a lesser extent by the structures. Figure 6.16C shows that this area is almost not occupied at all, when the structures of the compounds with a binding affinity higher than 100 nM (compounds **48**, **49**, **57** LY 288512, **60** LY 294920, **61** LY 294919, **31** L-364,718, **66** and **71**, the magenta coloured structures in Figure 6.16B) are left out. The remaining structures occupy similar positions as the best solutions found in the initial rigid docking trial with DRAGHOME and as the best solutions selected with the method based on the average of the weighted rankings of the four properties scores.

The second selection method, based on grouping the solutions on the basis of RMSDs of the non-hydrogen atoms of the ligand, requires the large number of solutions created by DRAGHOME using multiple conformations of a ligand. Therefore, this approach needs the results of a previous broad conformational search for each ligand. However, not only the conformational search and the docking of all obtained conformations with DRAGHOME, but also the selection method itself is computationally rather extensive for each ligand. Therefore the application of this procedure would not be useful for fast docking of a large number of different compounds. However, the method showed that a selection only on the total score of DRAGHOME could sometimes give not very plausible solutions, since sometimes solutions were found with a very low total score and with a very low number of hits, an indication that these solutions are not very favorable and rather extreme. Furthermore, the method showed that solutions with very similar orientations can have significant different total scores.
The first selection method on basis of the average of the weighted rankings of the four properties scores is computational less extensive, but it requires some hypothetical rules which can only be validated by experimental evidence such as mutagenesis data. The selection method based on grouping the many solutions of a ligand on the RMSD of the non-hydrogen atoms of the ligand, is computationally rather extensive, but probably produces less biased coverage of docking space. However, it loses most of its better coverage by somewhat arbitrary manual intervention in the selection of solutions.
Since the orientations of the best-ranked solutions selected by method one and method two are so close to each other and the second selection method which was applied to guarantee a better coverage of docking space, does not provide any advantage. Accordingly in the following only the orientations of the first approach are further considered.

6.3.4 Relaxation of Protein-Ligand Complexes

As described above, complexes of the solutions of the ligands with one of the conformations of the receptor are not ideal 3D-models due to frequent unfavourable contacts produced by the docking onto averaged representations of the receptor structure. Only complexes with a relaxed

structure without unfavourably close contacts can be analysed with a scoring function. Therefore, energy minimisations and/or molecular dynamic simulations were conducted for several complexes. Only favourable docking solutions were subjected to this simulation protocol (only 33 out of 48, see Table 6.7). The best solution, selected by applying the strategy of the average of the weighted rankings of the four properties and the three "urea-group"-rules, were transferred in their orientation into the representative conformations of the first five clusters of the clustering of the receptor conformations (Table 6.6). All these complexes (33 x 5) were relaxed with the basic minimisation protocol given in Table 6.4.
The root mean square deviations (RMS) of the non-hydrogen atoms of the ligand for all five relaxed complexes compared to the initial conformation of the ligand at the start, were calculated. The average and the standard deviation of these RMS values are depicted in Figure 6.17 as empty boxes. In the same Figure, also the average and the standard deviation of the RMS of the key atoms of the 14 important residue sidechains are depicted (Table 6.1).

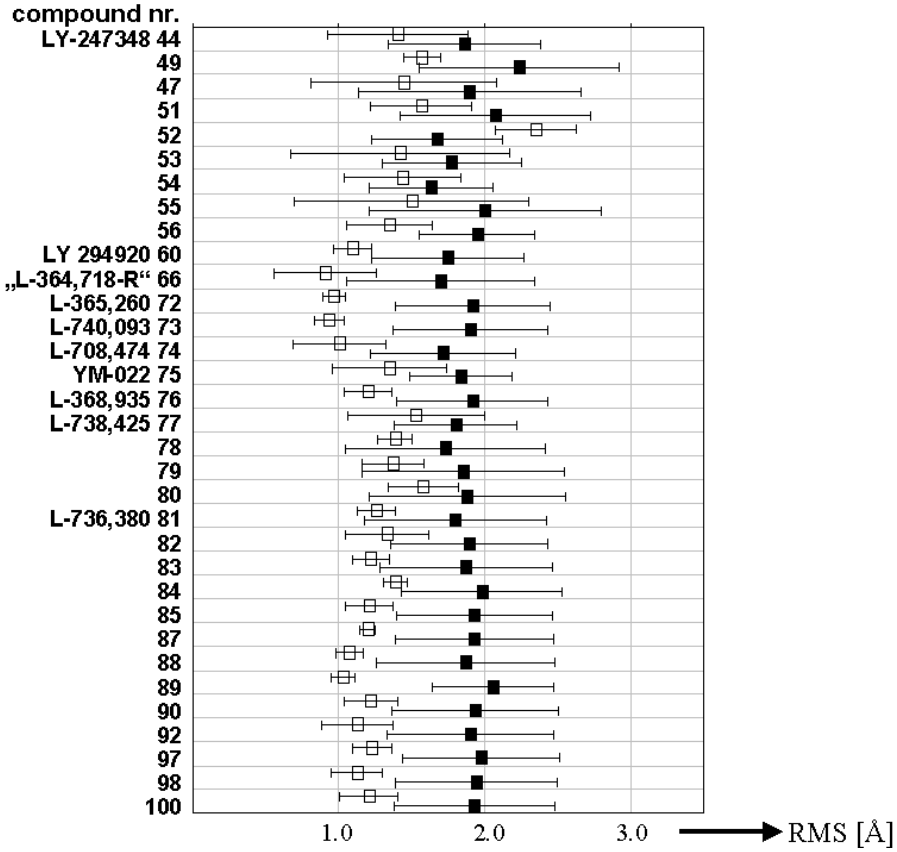


Figure 6.17: The average and standard deviation of the RMS of the heavy atoms of the ligands (empty box) and of the 14 atoms of the important residue sidechains of the receptor protein as listed in Table 6.1 (black box) in the 5 different minimised complexes for every computed compound.

Visual inspection reveals that the overall orientation of the ligands do not change significantly, which is also indicated by the small RMS values for the non-hydrogen atoms of the ligands. However, the side chains of the receptor display increased movements as indicated by the RMS values of the 14 atoms of the important residue sidechains of the receptor. The side chains are shifted to remove repulsive contacts between the ligand and the receptor.

An assessment of the degree of relaxation of the complexes can be obtained with the program PROBE of Word et al. [242]. PROBE allows to monitor the close contacts between the ligand and the receptor before and after the minimisation. It generates dots in space on the van der Waals surface of atoms which are in close proximity to atoms of the binding parties. In Figure 6.18, the close contacts for compound **76** L-368,935 are depicted. The number of unfavourable contacts (red dots) are reduced after minimisation.

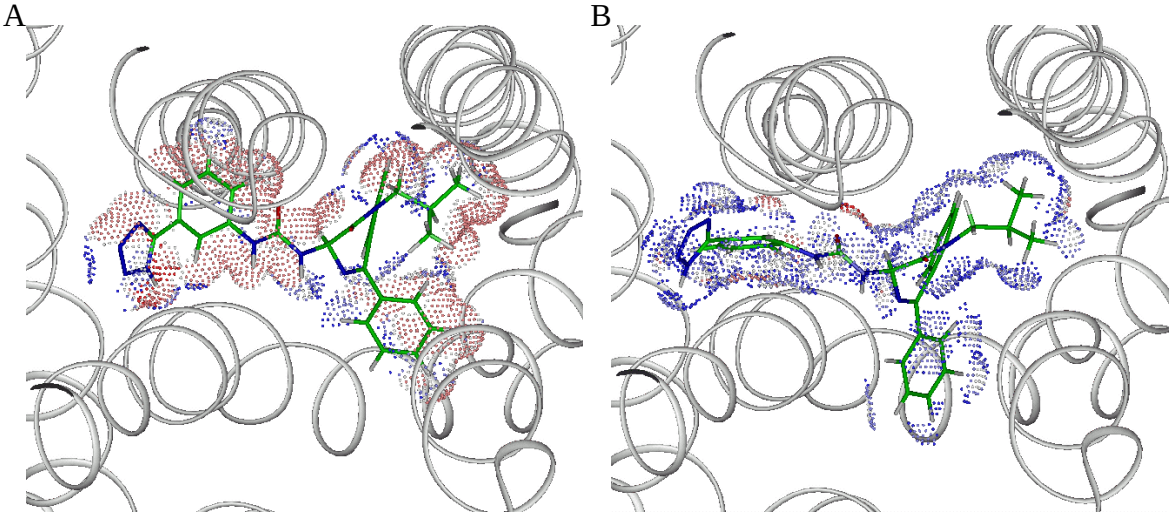


*Figure 6.18: Close contacts calculated by PROBE for compound **76** L-368,935, before minimisation (A) and after minimisation (B). The dots are colour-coded according to the contact distance from dark blue (far apart) over gray to red (overlap).*

Some quinazolinone derivatives are positioned after the docking procedure with part of their structures between the helices 5 and 6. As mentioned before, this is an unfavourable orientation, since the structures in this orientation would disrupt interhelical contacts between helices 5 and 6, and furthermore a structure gets into close contact to the lipid compartment. Minimisation of these quinazolinone derivatives does not remove these unfavourable contacts instead some ligands moved further towards the exterior of the receptor. This indicates poor docking results. Supposedly the docking and binding results of the quinazolinone derivatives are thus rather speculative.
Because the applied energy minimisations were rudimentary, not all the complexes were fully relaxed and some unfavourable close contacts could still remain between ligand and receptor. In order to determine whether more elaborate protocols including molecular dynamic simulations are required, more strigent minimisations with subsequent molecular dynamic simulations were applied for a few selected complexes. For this purpose, frame 393 of the molecular dynamics run of Chapter 4 was used as receptor model, since it is the most representative conformation of the 14 clusters of the receptor (see Table 6.6). For compounds **50**, **53**, **57**, **66**, **73**, **75**, **80**, **81**, **84**, and **89** the extended relaxation script given in Table 6.5 was applied.
In the distance matrices of Figure 6.19, the RMS-values of the non-hydrogen atoms of the ligand (A), the key atoms of 14 amino acid residues of the protein (B), and the backbone atoms of the protein (C) for the complex of compound **89** recorded over the complete simulation time are depicted. The RMS-values are calculated with respect to the start conformation of the complex. It can be seen that the non-hydrogen atoms of the ligand do not change significantly (RMS < 1.0 Å). This picture is seen for all the complexes and indicates clearly that the initial position

of the ligand is not dramatically changed during the calculation. The RMS-values for all the backbone atoms of the protein and the key atoms of the 14 active site residues were significantly larger. However, also these values are still reasonably small and indicate that the structures of the complexes do not change dramatically. It was determined that the change of the backbone atoms is mainly determined by the changes of the residues in the loop regions.

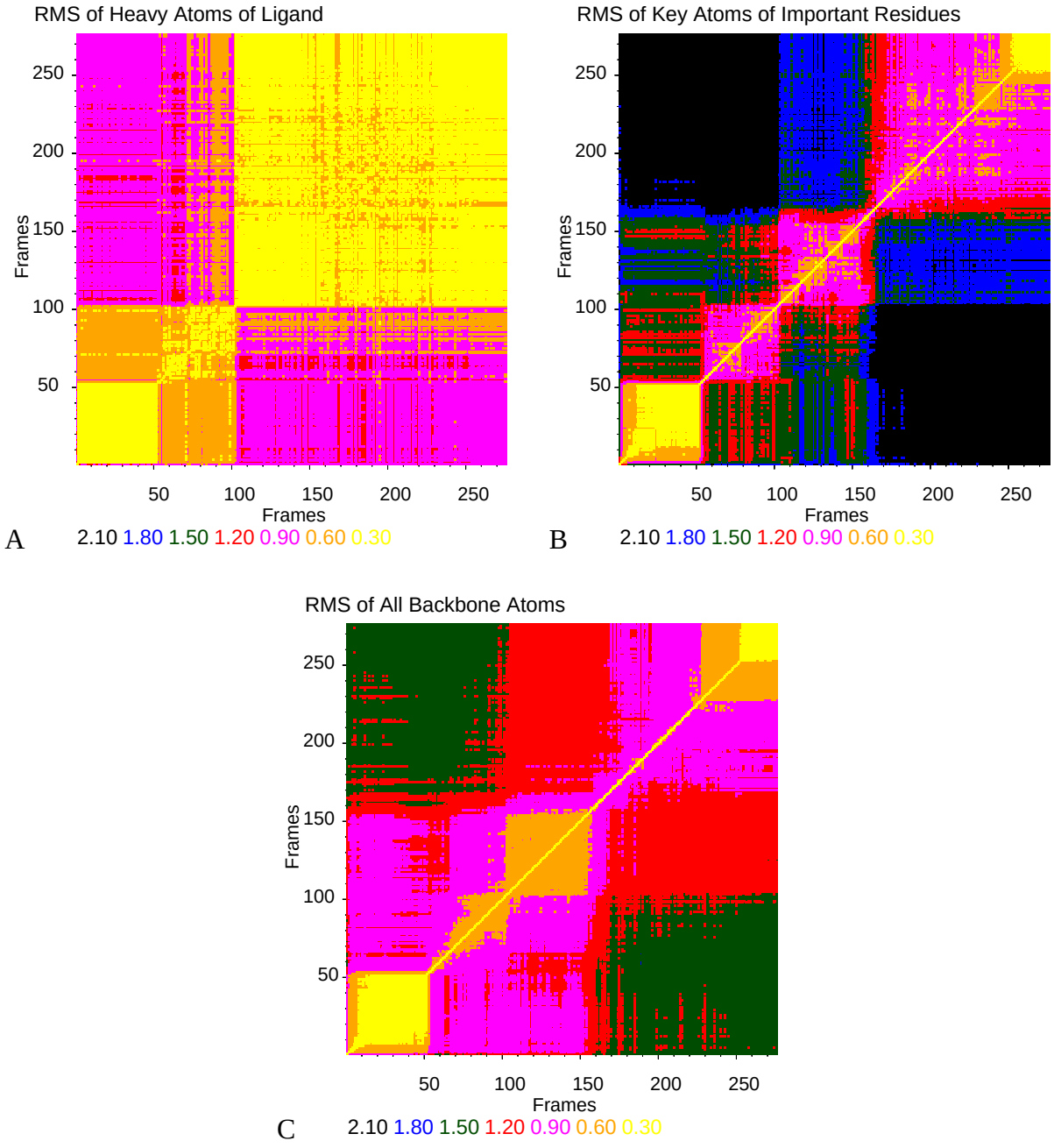


*Figure 6.19: RMS distance matrices of the complex of compound **89**, (A) the heavy atoms of the ligand, (B) the key atoms of the 14 for antagonist binding important residues of the receptor, and (C) all the heavy backbone atoms of the protein receptor.*

6.3.5 Putative Ligand Binding Site

Interacting residues within the different minimised complexes of the different ligands were analysed with the program PROBE and by visual inspection. From these analyses a schematic active site with the interacting amino acids can be developed (Figure 6.20). It is remarkable

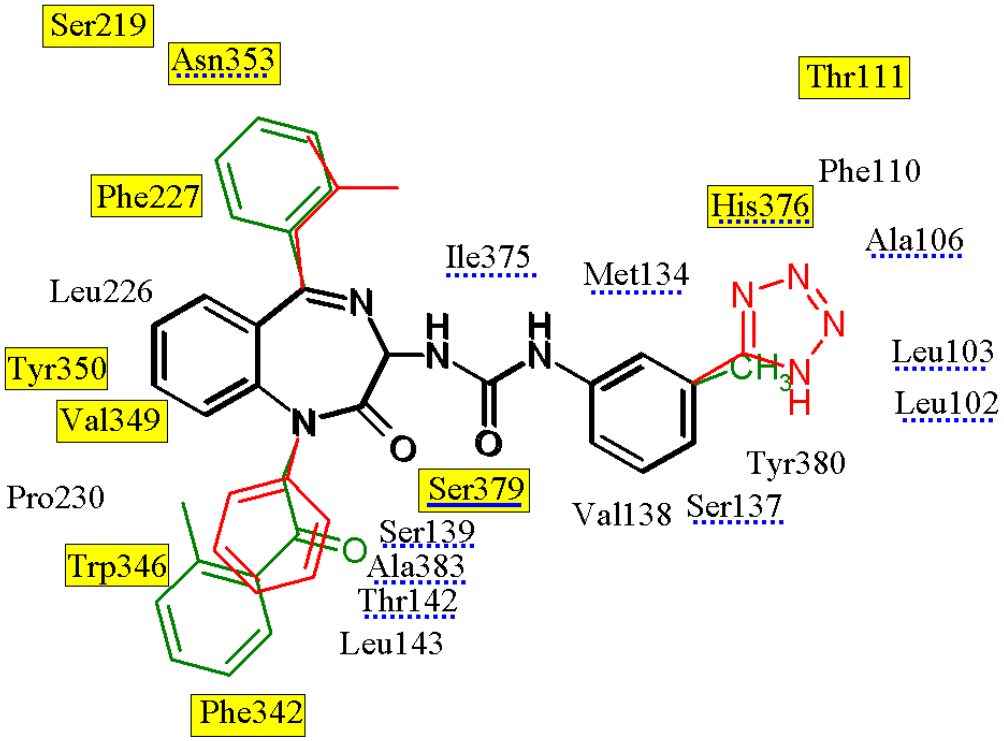


*Figure 6.20: Schematic view of the active site of the human CCK-B receptor with two benzodiazepine derivatives, notably compounds **76** (L-368,935) (red) and **75** (YM-022) (green). The interacting amino acids identified by several mutagenesis studies as important for antagonists binding are boxed, amino acids involved in >90% of the different complexes in hydrogen bonding are underlined, and amino acids occassionally involved in hydrogen bonding are underlined with a dotted line, other residues form hydrophobic interactions.*

that all amino acids discovered by several mutagenesis studies [147, 148, 149, 150, 151, 152, 153, 154] as being involved in non-peptide antagonist binding are identified in these analyses as interacting amino acids. However, also several other amino acids are interacting with the ligands. These amino acids are considered as interesting candidates for further mutagenesis. However, since some of these amino acids are interacting via their backbone atoms, mutagenesis of these residues studies will probably not always reveal conclusive results. So have Kopin et al. [147] already mutated Ala106 to Phe, but this mutation had no affect on non-peptide antagonist binding.
Furthermore, not all ligands interact via the same pattern. Already some benzodiazepine derivatives with similar sidechains are indicated to have different interaction patterns and/or partners. This holds especially for hydrophobic interactions. Since hydrophobic interactions are very difficult to quantify, it is difficult to indicate in an unambiguous way the interactions between the amino acid residues and different ligands.

6.3.6 CoMSIA Analyses

A CoMSIA analysis was performed considering two ligand alignments. The first alignment was obtained by using the best solution of the second flexible docking trial with DRAGHOME selected by the strategy on basis of the average of the weighted rankings of the four properties including the "urea-group"-rule. The second alignment was obtained by selecting visually for each ligand of the first alignment the best conformation of the five minimised conformations of each ligand (see also section 6.2.4) on basis of the mutual spatial similarity strategy of Schafferhans [221].

CoMSIA reveals statistical significant models for both alignments, however these models should be approached with some caution, since the number of outliers is rather high. The statistical data is shown in Table 6.8.

	CoMSIA start-conf.	CoMSIA end-conf.
Statistics		
q^2	0.512	0.500
s_{PRESS}	0.855	0.745
r^2	0.943	0.951
S	0.292	0.233
# comp.	4	3
Box		
Stepsize (Å)	1.0	1.0
x	-12 to +18	-12 to +18
y	-13 to +8	-13 to +8
z	-13 to +11	-13 to +11
# points	17050	17050
Fraction		
steric	0.132	0.111
electrostatic	0.245	0.228
hydrophobic	0.250	0.249
H-donor	0.195	0.222
H-acceptor	0.110	0.190
# ligands	29	24
# outliers	4	9

Table 6.8: Statistics of CoMSIA models of the ligand alignments before and after minimisation calculations.

In Figures 6.21 and 6.22, the different contribution maps of the CoMSIA analysis are depicted for compounds **51**, **75**, **76** (L-368,935), and **77** (L-738,425) as top-view and as side-view, respectively. Very similar results were obtained for the CoMSIA analyses of both alignments, accordingly only contribution maps of the second alignment are given. The contribution maps are contoured at 90% and 10% level and correspond to the following colour scheme:

property field	colour 90% favored level	colour 10% disfavored level
steric	green	orange
hydrophobic	orange	blue
electrostatic	green	red
donor	red	magenta
acceptor	blue	purple

As expected, contribution maps of donors and acceptors are found next to the urea-group in both models. A clearly disfavoured volume is found between helices 5 and 6 (orange and blue contour maps, disfavored steric and hydrophobic maps, repectively, in Figures 6.21A, B and 6.22A, B). It was already mentioned before that this area is not a favourable area for the ligand binding due to too close proximity of the lipid compartment. Favourable steric interactions are predicted near the phenyl of the urea-substituent and at the core of the benzodiazepine. Furthermore, one ring-substituent is positioned in a favourable hydrophobic area (orange area in Figures 6.21B and 6.22B). The ring-substituents do not indicate any favourable or disfavourable contribution of donor, acceptor or electrostatic properties in close proximity (Figures 6.21C, D, E and 6.22C, D, E).
The binding affinities for the ligands predicted with both CoMSIA models are listed in Table 6.9, together with the delta to the experimental bindings energies and the delta between the predicted binding energies of the two models.

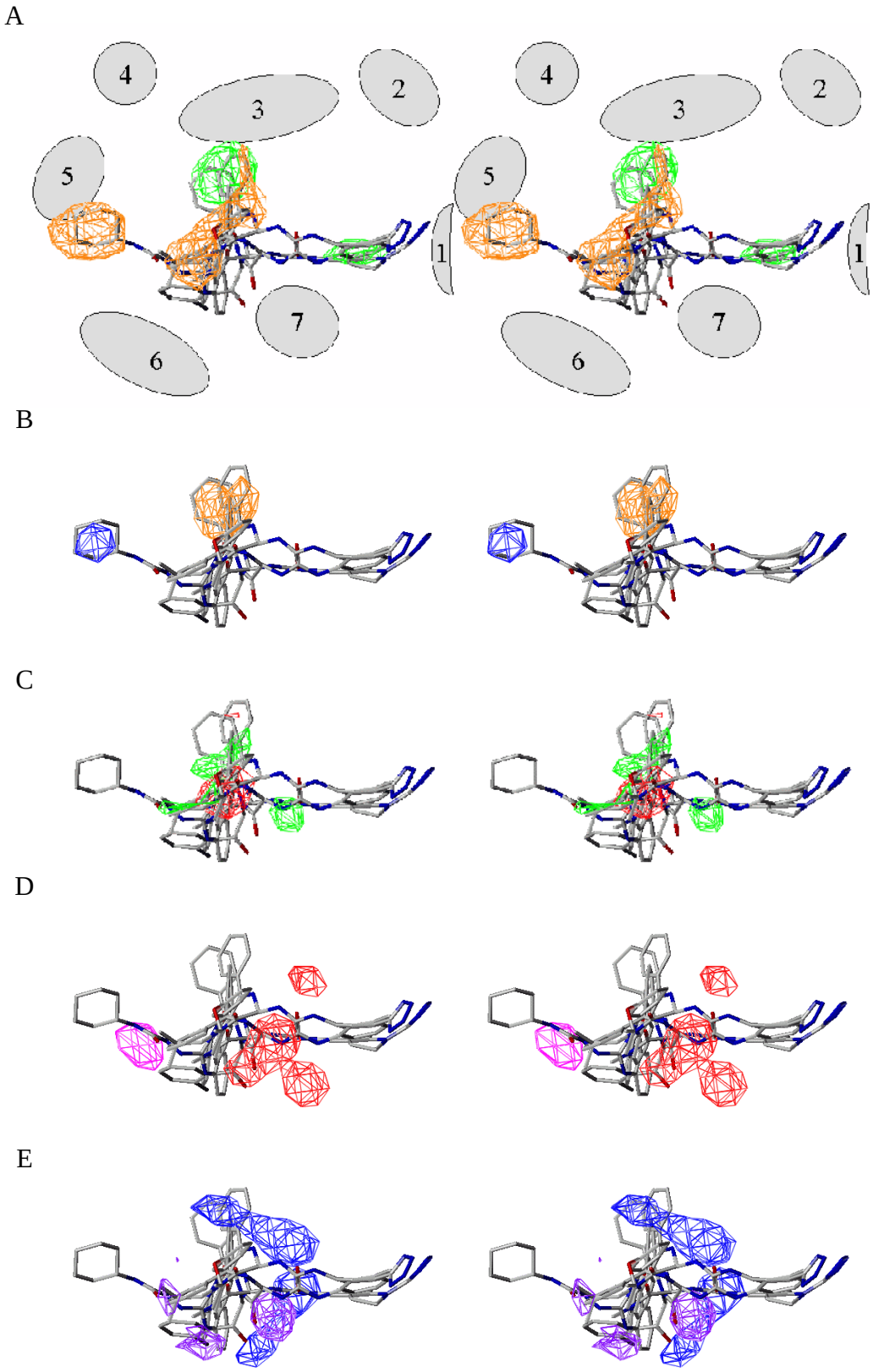


Figure 6.21: Stereo topview of the property fields of the CoMSIA-analysis of the ligand alignment after the minimisation calculation of the complexes, notably (A) steric, (B) hydrophobic, (C) electrostatic, (D) donor, and (E) acceptor. In (A) are the positions of the helices schematic depicted by gray ovals. See text for more details.

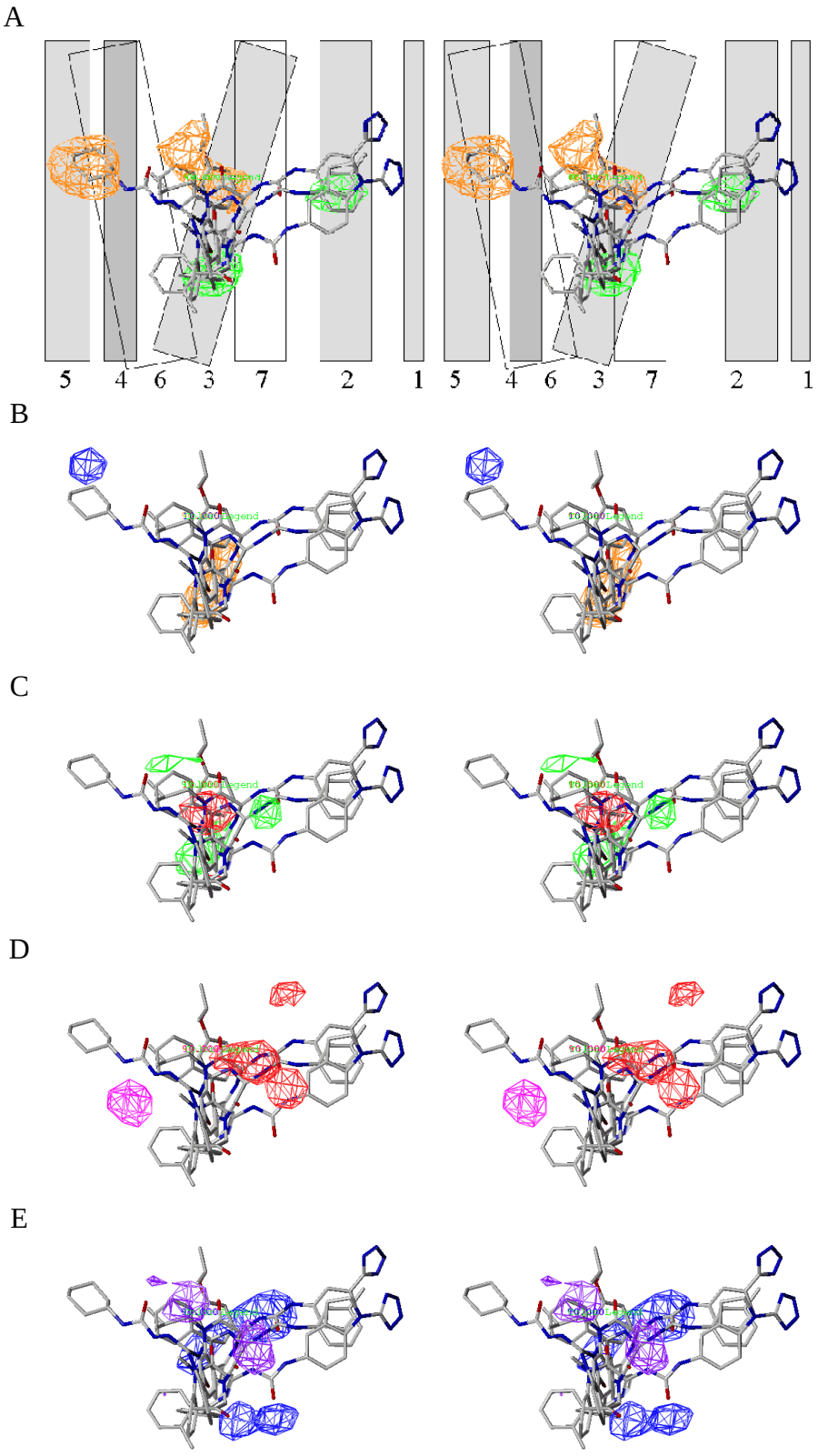


Figure 6.22: Stereo sideview of the property fields of the CoMSIA-analysis of the ligand alignment before the minimisation calculation of the complexes, notably (A) steric, (B) hydrophobic, (C) electrostatic, (D) donor, and (E) acceptor. In (A) are the positions of the helices schematic depicted by rectangulars. See text for more details.

compound name	IC50	log IC50	predicted log IC50 start conf.	delta log IC50 start - exp.	prediction logIC50 end conf.	delta logIC50 end - exp.	delta logIC50 start conf.- end conf.
66 L-364,718 R	3700.00	5.43	9.06	3.63	9.58	4.15	0.52
72 L-365,260	2.00	8.70	8.60	-0.10	8.72	0.02	0.13
76 L-368,935	0.14	9.85	9.54	-0.32	9.29	-0.56	-0.25
74 L-708,474	0.28	9.55	9.63	0.08	9.72	0.16	0.09
82	137.00	6.86	7.19	0.33	8.77	1.91	1.58
81 L-736,380	0.05	10.27	9.95	-0.32	8.77	-1.5	-1.18
77 L-738,425	0.11	9.96	9.58	-0.38	10.07	0.11	0.49
73 L-740,093	0.10	10.00	8.29	-1.71	8.51	-1.49	0.23
44 LY 247348	26.00	7.59	7.81	0.22	8.99	1.41	1.18
83	1.50	8.82	9.02	0.20	8.59	-0.23	-0.43
84	54.00	7.27	7.61	0.35	7.43	0.17	-0.18
87	480.00	6.32	6.97	0.65	8.09	1.77	1.12
88	23.00	7.64	7.19	-0.45	8.00	0.36	0.82
89	8.00	8.10	8.14	0.04	8.03	-0.06	-0.11
92	84.00	7.08	7.13	0.06	7.36	0.29	0.23
97	1.00	9.00	7.53	-1.47	7.84	-1.16	0.31
100	26.00	7.59	7.19	-0.39	7.79	0.21	0.60
48	585.00	6.23	6.35	0.11	8.71	2.47	2.36
50	14.00	7.85	7.57	-0.28	7.69	-0.16	0.12
55	97.00	7.01	6.94	-0.08	7.04	0.03	0.10
56	63.00	7.20	7.11	-0.09	6.92	-0.28	-0.20
52	29.00	7.54	7.64	0.10	7.52	-0.02	-0.12
53	1.00	9.00	9.25	0.25	9.09	0.09	-0.16
79	7.50	8.12	8.18	0.05	8.07	-0.05	-0.11
78	0.58	9.24	9.11	-0.12	9.15	-0.09	0.03
80	1.10	8.96	9.26	0.30	8.91	-0.05	-0.35
51	108.00	6.97	6.92	-0.04	6.95	-0.02	0.02
54	11.00	7.96	7.75	-0.21	7.96	0.01	0.21
85	18.00	7.74	7.44	-0.31	8.03	0.29	0.60
98	48.00	7.32	7.20	-0.12	7.29	-0.03	0.09
90	2.00	8.70	8.80	0.10	8.33	-0.37	-0.47
60	1900.00	5.72	8.37	2.65	8.58	2.86	0.21
75 YM022	0.11	9.96	10.32	0.36	10.165	0.21	-0.16

Table 6.9: Analysis of the two CoMSIA models. The two predicted binding energies, the deltas between experimental and predicted binding energies, and in the last column the delta between the two predicted binding energies are given for 33 compounds.

6.3.7 Docking of CCK-4 and CCK-8

The docking procedure applied to the 48 antagonists was also used to generate complex structures with CCK-4 and CCK-8 as native ligands. In Figures 6.23 and 6.24, the top 10 solutions for CCK-8 and CCK-4 docking do not achieve corresponding binding modes. They are positioned in anti-parallel fashion in the receptor. However, they occupy the same compartment in the receptor, namely the cavities one and two. These are the cavities located between helices 3, 4, 5, 6 and 7 at two different levels within the transmembrane domain of the receptor. CCK-4 and CCK-8 do not make contacts to cavity three in the proximity to His376.
When these results are compared to experimental findings of the binding of lipo-tetragastrin derivatives [243, 244, 120] and peptide/benzodiazepine hybrids [245], carried out by the group

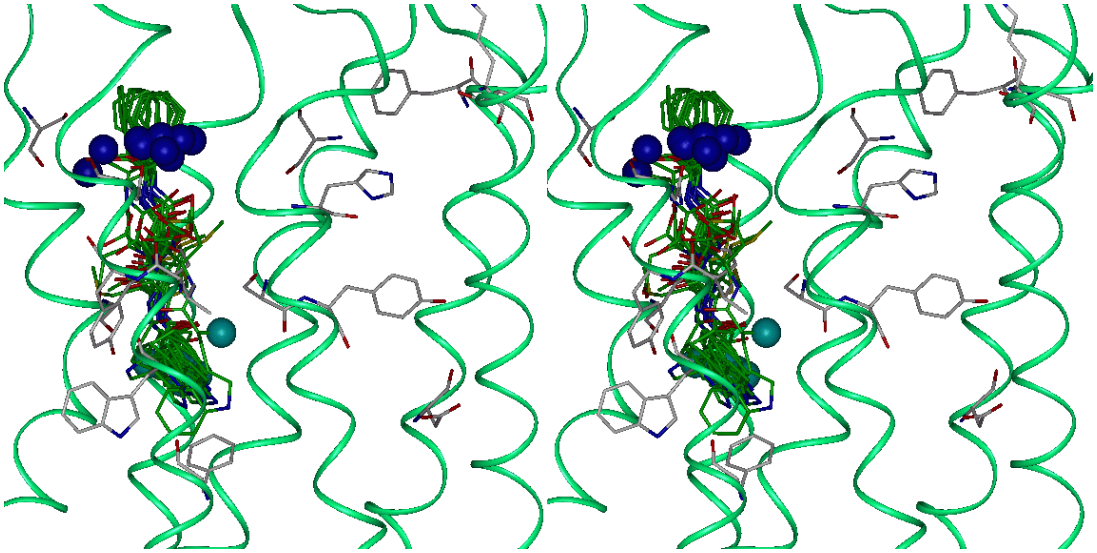


Figure 6.23: Side-by-side stereo presentation of the first 10 solutions of CCK-4 in the receptor. The N-atom of the NH_2-group of Phe (the head) is a CPK with dark blue colour and the N-atom of the backbone of Trp (the tail) is a CPK coloured light blue.

of L. Moroder, the simulations in this study are in accordance with their results. The agonists bind to the same binding area, but with different binding patterns. Also the kink, mentioned by Lutz [120] between Ala and Tyr in the sequence of Gastrin (H-Pyr-Gly-Pro-Trp-Leu-$(Glu)_5$-Ala-Tyr-Met-Gly-Trp-Met-Asp-Phe-NH_2) was also observed for Asp and Tyr in the sequence of CCK-8 (H-Asp-Tyr(SO_3H)-Met-Gly-Trp-Met-Asp-Phe-NH_2). Furthermore, CCK-4 and CCK-8 are involved in interactions with amino acids indicated by mutagenesis studies to be important for agonist binding, as, e.g., His207, Arg 208, Trp209, Asn353, and Trp346. However, interactions with, e.g., Thr111, Asn115, Leu116, Phe120 and Phe122 were not observed. These last group of residues are positioned in the extracellular loop 2 and this loop is bend away from the top of the receptor. Therefore, these amino acids are in our model too far apart, but could possible be much closer in space when the loop is bend towards the top of the receptor. This might be attributed to the large flexibility of the loop regions of the receptor and of the ligand, notably CCK-8. The approximative character of this study does not allow to conclude on detailed interaction schemes, only a general idea can be derived in a highly qualitative manner.

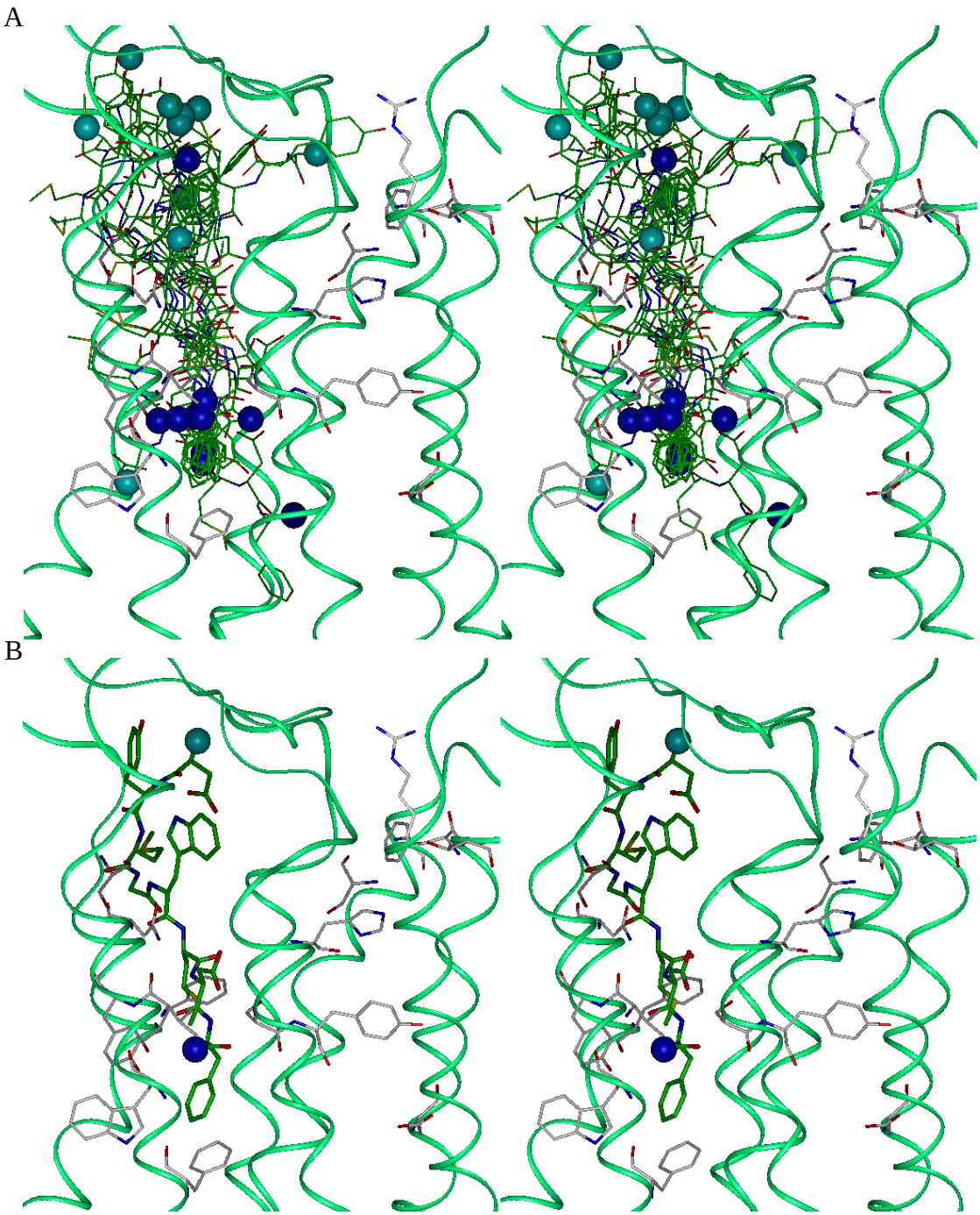


Figure 6.24: Side-by-side stereo presentation of the first 10 docking solutions of CCK-8 (A) and a representative orientation of CCK-8 in the receptor. The N-atom of the NH_2-group of Phe (the head) is a CPK with dark blue colour and the N-atom of the backbone of Asp (the tail) is a CPK coloured light blue.

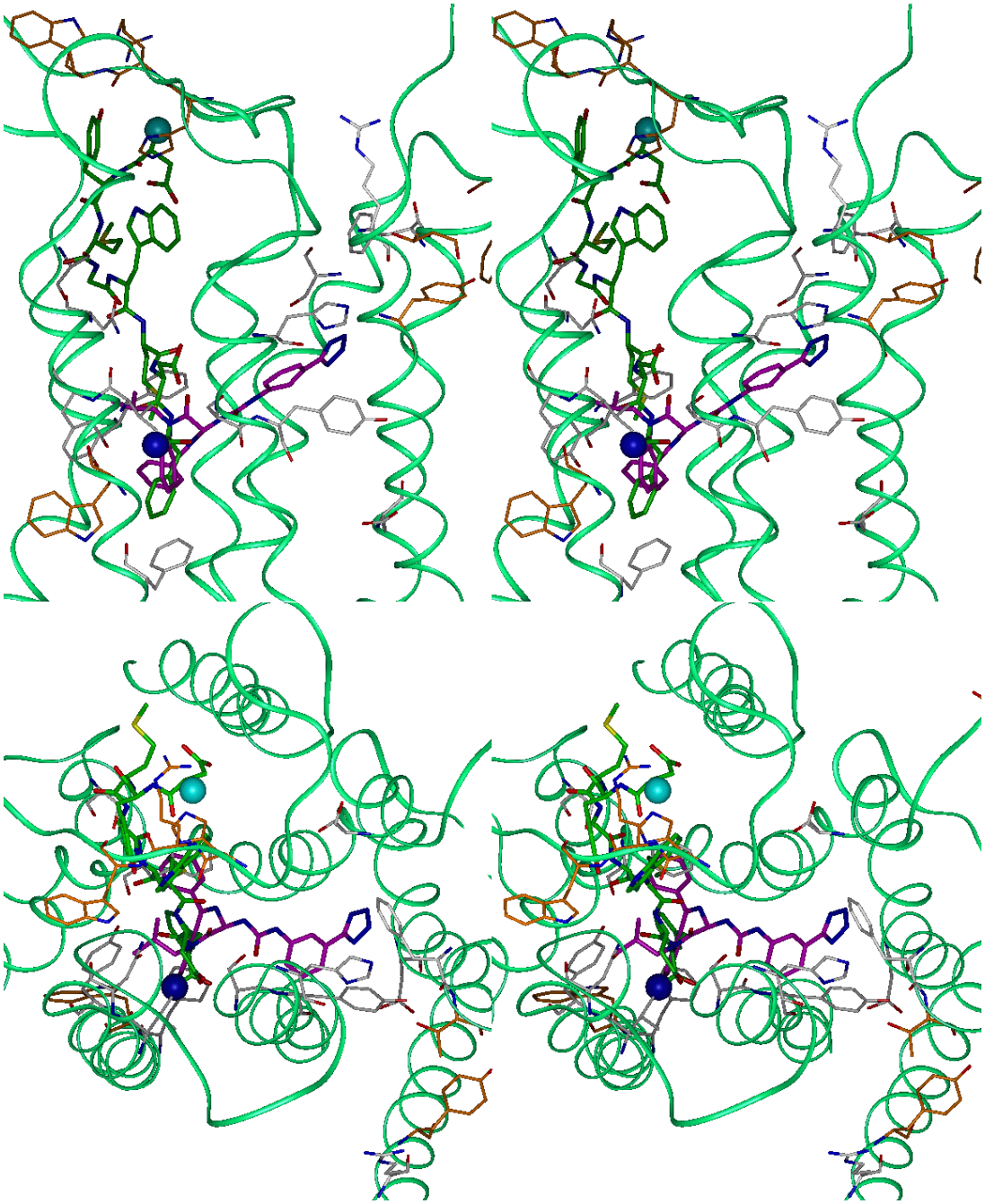


Figure 6.25: Side-by-side stereo presentation of CCK-8 (green) with L-368,935 (magenta) docked within the receptor. The orange coloured amino acids are important for binding of agonists, while the gray coloured amino acids are only important for antagonist binding.

6.4 Conclusion

It can be concluded that the benzodiazepine derivatives could be docked into the CCK-B receptor model with DRAGHOME in an orientation in close proximity to amino-acid residues previously identified as being involved in antagonist binding. In Figure 6.20, the binding site for these benzodiazepine derivatives is depicted schematically. The urea group is located between helices 3 and 7. It is capable to form hydrogen bonds to Ser379 and has a further hydrogen bonding partner in either Ile374 or Met134. Ser379 at helix 7 was indicated by mutagenesis to play an important role in the binding of non-peptide CCK-B antagonists. Furthermore, it should be noted that helices 3 and 7 also comprise amino-acid residues which play a key role in ligand binding with other GPCRs. The important lysine residue of the rhodopsin receptors is located in helix 7 and the aspartate of the mono-amine receptors is located in helix 3, in close proximity to Met134 and Ser137 of the CCK-B receptor.
The two ring substituents of the benzodiazepine derivatives are positioned into a large hydrophobic pocket. This hydrophobic pocket was clearly detected by the DRAGHOME approach. Several aromatic residues border this pocket and make contacts to the ligand. It could not be determined whether there is a specific preference for the placement of the two ring substituents of the asymmetric benzodiazepine derivatives within this pocket. It might well differ from ligand to ligand, since Asn353 in the upper part and Ser382 and Thr142 in the lower part of this pocket are likely polar groups to possibly anchor these ring substituents. However, these three residues are not the only possible contacting groups, also the backbone atoms and several other residues are putative candidates. In addition, the hydrophobic residues can take influence on the adopted orientation of the individual ligands.
In summary, the advantages of the applied docking procedure should be highlighted:

- DRAGHOME accounts for protein flexibility and limited accuracy of the receptor protein model
- MOCCA analysis of the ligand accounts for the flexibility of the ligand
- combining protein flexibility with ligand flexibility accounts for an exhausive sampling of the docking space of a ligand, however scoring of the produced solutions becomes an increasing problem
- orientation of ligands within the protein model is achieved without predescribed binding contrains and/or interactions
- no overestimation of hydrogen bonds, the contribution of steric parameters is important
- even small cavities deep within the protein receptor are identified
- new amino acids possibly involved in non-peptide antagonists binding are identified
- a general idea was developed for the binding of CCK-8

In general can be concluded that the DRAGHOME docking procedure obtains results which can be used to plan and suggest subsequent experimental studies. The obtained binding geometries are clearly not precise enough to allow for a meaningful calculation of binding affinities.

Chapter 7

Summary, Zusammenfassung

"When I am judging a theory, I ask myself whether, if I were God, I would have arranged the world in such a way."

- Albert Einstein
  (**1879-1955**), scientist and Nobel Prize in Physics laureate

7.1 Summary

Even though pharmaceutical research has entered the frequently quoted post-genome era, it is still undisputed that G protein-coupled receptors constitute the most densely populated protein family encompassing numerous disease-relevant drug targets. Consequently, medicinal chemistry is expected to pursue targets from that protein family in that hits need to be generated and subsequently optimized towards viable clinical candidates for a variety of therapeutic areas. For the purpose of rationalizing structure-activity relationships within such optimization programs, structural information derived from the ligand's as well as the macromolecule's perspective is essential. While it is relatively straightforward to define pharmacophore hypotheses based on comparative modelling of structurally and biologically characterized low-molecular weight ligands, a deeper understanding of the molecular recognition event underlying, e.g., an antagonist-GPCR interaction remains challenging, since the principally available amount of experimentally derived structural data on GPCRs is extremely scarse when compared to, e.g., soluble enzymes.

In this context, the protein modelling methodologies introduced, developed, optimized, and applied in this thesis provide structural models that are capable of assisting in the development of structural hypotheses on ligand-receptor complexes. As such they provide a valuable structural framework not only for a more detailed insight into ligand-GPCR interaction, but also for guiding the design process towards next-generation compounds which should display enhanced affinity.

The model building procedure developed in this thesis systematically follows a hierarchical approach, sequentially generating a 1D topology, followed by a 2D topology that is finally converted into a 3D topology.

The determination of a 1D topology is based on a compartmentalization of the linear amino acid sequence of a GPCR of interest into the extracellular, intracellular, and transmembrane sequence stretches. The entire chapter 3 of this study elaborates on the strengths and weaknesses of applying automated prediction tools based on the calculation of property profiles, distribution profiles, and property moments for the purpose of identifying the transmembrane sequence domains. Based on an once derived 1D topology, a type of in-plane projection structure for the seven transmembrane helices can be derived with the aide of calculated vectorial property moments, yielding in the 2D topology. Thorough bioinformatics studies revealed that only a consensus approach based on a conceptual combination of different methods employing a carefully made selection of parameter sets gave reliable results, thus emphasizing the danger of any attempt to fully automate a GPCR modelling procedure.

Chapter 4 of this thesis describes a tailored procedure to further expand the 2D topological findings into 3D space, exemplified on the human CCK-B receptor protein. This particular GPCR was chosen as the receptor of interest, since an enormous experimentally derived and structurally relevant data-set was available, ranging from detailed structure-activity relationships of low-molecular weight agonists and antagonists (outlined in detail in chapter 5), to mutagenesis data that identified receptor protein residues putatively involved in ligand binding. Within the computational refinement procedure of constructed GPCR models, major emphasis was laid on the explicit treatment of a non-isotropic solvent environment during molecular mechanics (i.e. energy minimization and molecular dynamics simulations) calculations. The majority of simulations was therefore carried out in a tri-phasic solvent box accounting for a central lipid

environment, flanked by two aqueous compartments, mimicking the extracellular and cytoplasmic space.

Chapter 5 introduces the reference compound set, comprising low-molecular weight compounds modulating CCK receptors, that was used for validation purposes of the generated models of the receptor protein.

Chapter 6 describes how the generated model of the CCK-B receptor was subjected to intensive docking studies employing compound series introduced in chapter 5. It turned out that by applying the DRAGHOME methodology viable structural hypotheses on putative receptor-ligand complexes could be generated. Based on the methodology pursued in this thesis a detailed model of the receptor binding site could be devised that accounts for known structure-activity relationships as well as for results obtained by site-directed mutagenesis studies in a qualitative manner.

The overall study presented in this thesis is primarily aimed to deliver a feasibility study on generating model structures of GPCRs by a conceptual combination of tailor-made bioinformatics techniques with the toolbox of protein modelling, exemplified on the human CCK-B receptor.

In summary, it is again emphasized that the generated structures should be envisioned as models only, not necessarily providing a detailed image of reality. However, consistent models, when verified and refined against experimental data, deliver an extremely useful structural contextual platform on which different scientific disciplines such as medicinal chemistry, molecular biology, and biophysics can effectively communicate.

7.2 Zusammenfassung

Obgleich die pharmazeutisch-chemische Forschung das postgenome Zeitalter erreicht hat, repräsentiert die Familie der G Protein-gekoppelten Rezeptoren zweifellos eine der höchst-populierten Proteinfamilien, die zahllose Krankheits-relevante Zielproteine enthält. Folglich wird die Medizinische Chemie Repräsentanten dieser Rezeptorfamilie bearbeiten, d. h. Hit-Strukturen generieren und diese zu vielversprechenden Kandidaten für die klinische Entwicklung weiter optimieren, wobei sich ein breites Spektrum therapeutischer Indikationen adressieren lässt. Zum tiefergehenden Verständnis von Struktur-Wirkungsbeziehungen im Kontext solcher Optimierungsprogramme liefern Strukturinformation über Liganden und über die Rezeptorproteine wertvolle Beiträge. Während Pharmakophorhypothesen relativ einfach aus vergleichenden Studien von strukturell wie biologisch charakterisierten niedermolekularen Substanzen aufzustellen sind, stellt die Generierung eines detaillierten Verständnisses über das zugrundeliegende molekularen Erkennungsgeschehen von z. B. Antagonist-GPCR Wechselwirkung nach wie vor eine grosse Herausforderung dar. Dies liegt unter anderem daran, dass die prinzipiell zur Verfügung stehende, experimentell abgeleitete Strukturinformation im Vergleich zu z. B. löslichen Enzymen extrem limitiert ist.

Vor diesem Hintergrund erlauben die in dieser Arbeit eingeführten, entwickelten und optimierten Methoden der Proteinmodellierung die Erstellung von Strukturmodellen, die wertvolle Hilfestellung zum Aufstellen von Strukturhypothesen zu Ligand-Rezeptorkomplexen liefern. Dadurch wird es nicht nur möglich, Ligand-Rezeptor Interaktion auf einem detaillierteren Niveau zu verstehen, es wird vielmehr auch der Designprozess zu Verbindungen der nächsten Generation beeinflusst, der letztlich zu Substanzen mit verbesserter Affinität führen soll.

Das in dieser Arbeit entwickelte Verfahren zum Modellbau von Rezeptorproteinen folgt systematisch einem hierarchischen Vorgehen, das sequentiell eine 1D Topologie erzeugt, gefolgt von einer 2D Topologie die letztendlich in eine 3D Topologie transformiert wird.
Die Erstellung der 1D Topologie beruht auf einer Kompartmentierung der linearen Primärstruktur, sprich der Aminosäuresequenz des GPCRs in extrazelluläre, intrazelluläre und transmembrane Sequenzabschnitte. Kapitel 3 dieser Arbeit setzt sich intensiv mit den Vor- und Nachteilen der Anwendung automatisierter Vorhersage Programme auseinander, die im wesentlichen auf der Berechnung von Eigenschaftsprofilen, Verteilungsprofilen und vektoriellen Eigenschaftsmomenten beruhen, anhand derer die potentiell transmembranen Sequenzbereiche identifizierbar sein sollten.
Ausgehend von einer einmal erstellten 1D Topologie wird eine Art Projektionsstruktur für die sieben transmembranen Helices erstellt, wobei die Projektionsebene parallel zur Membranoberfläche liegt. Diese Projektionsstruktur (2D Topologie) wird mit Hilfe der vektoriellen Eigenschaftsmomente der einzelnen Helices erhalten. Intensive Studien mit den Werkzeugen der Bioinformatik führten zu dem Resultat, dass nur ein Konsensus-Vorgehen, also eine konzeptionelle Kombination unterschiedlicher Methoden, basierend auf sorgfältig ausgewählten Parametersätzen valide 2D Topologien lieferte, was auf die Gefahr hinweist, die sich in der Anwendung voll-automatisierter GPCR Modellingverfahren zur Modellgenerierung verbirgt.
Kapitel 4 der vorliegenden Arbeit beschreibt ein maßgeschneidertes Procedere zur Transformation der bis dahin abgeleiteten 2D Topologie in ein 3D Strukturmodell, was exemplarisch am humanen CCK-B Rezeptorprotein demonstriert wird. Dieser spezielle Rezeptor wurde ausgewählt, da eine Grosszahl experimentell abgeleiteter strukturrelevanter Daten vorhanden sind, die von detaillierten Struktur-Wirkungsbeziehungen niedermolekularer Agonisten und Antagonisten (siehe Kapitel 5) bis zu Mutagenese-Studien zur Identifikation Ligand-bindender Rezeptorprotein-Aminosäuren reichen.
Zur Computer-gestützten energetischen Verfeinerung konstruierter GPCR-Modelle wurde die Hauptaufmerksamkeit auf die explizite Berücksichtigung der anisotropen Umgebung im Rahmen von molekulardynamischen-Simulationen (Energieminimierung und Molekulardynamik Rechnungen) gelegt. Annähernd alle Simulationen wurden demzufolge in einem speziell entwickelten Drei-Phasen-System durchgeführt, das aus einer zentralen Lipidphase besteht, die von zwei wässrigen Phasen flankiert wird, welche das extrazelluläre und das cytoplasmatische Kompartiment imitieren sollen.
Im Kapitel 5 wird der Ligandsatz vorgestellt, der als Referenz zur Validierung der Rezeptormodelle im Zuge der in Kapitel 6 beschriebenen Docking-Studien darstellt. Im Kontext dieser Docking-Studien stellte sich heraus, dass durch Verwendung der DRAGHOME Methodik validierbare Strukturmodelle zur Rezeptor-Ligand Interaktion erhalten werden konnten.
Mithilfe des in dieser Studie verwendeten und entwickelten Methodenarsenals gelang es schließlich, ein detailliertes Modell der Rezeptor-Bindungstasche zu formulieren, das sowohl bekannte Struktur-Wirkungsbeziehungen, als auch Ergebnisse aus Mutagenese-Studien zufriedenstellend zu erklären vermag. Ein wesentliches Ziel der Arbeit bestand auch darin, eine Art Machbarkeitsstudie zur Modellierung von GPCRs durchzuführen, indem konzeptionell maßgeschneiderte Bioinformatik-Techniken mit Methoden der Proteinmodellierung kombiniert wurden, was am Beispiel des humanen CCK-B Rezeptors demonstriert worden ist.
Zusammenfassend sei an dieser Stelle erneut hervorgehoben, dass die erzeugten Strukturmodelle auch wirklich als Modelle einzustufen sind, die nicht notwendigerweise mit den realen Gegebenheiten am Rezeptor übereinstimmen. Trotzdem stellt ein in sich konsistentes Mod-

ell, sofern es ausreichend verifiziert und gegen experimentelle Daten verfeinert ist, eine extrem hilfreiche Plattform dar, auf der unterschiedlichste wissenschaftliche Disziplinen miteinander effektiv kommunizieren können, was die Medizinische Chemie, die Molekularbiologie und die Biophysik betrifft.

Chapter 8

References

"It is seldom that authors report on experiments that yield disappointing results and such reports are to be encouraged, because new ideas are often precipitated in such attempts."

- J. Zupan and J. Gasteiger
  Anal.Chim.Acta (**1991**) 248: 1-30

[1] Ockey, D. A & Gadek, T. R. Inhibitors of protein-protein interaction. *Exp. Opin. Therap. Patents* (**2002**) 12, 393–400.

[2] Toogood, P. L. Inhibition of protein-protein association by small molecules: approaches and progress. *J. Med. Chem.* (**2001**) 45, 1543–1558.

[3] Doyle, D. A, cabral, J. M, Pfuetzner, R. A, Kuo, A, Gulbis, J. M, Cohen, S. L, Chait, B. T, & MacKinnon, R. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* (**1998**) 280, 69–77.

[4] Xiong, J. P, Stehle, T, Diefenbach, B, Zhang, R, Dunker, R, Scott, D. L, Joachimiak, A, Goodman, S. L, & Arnaout, M. A. Crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$. *Science* (**2002**) 294, 339–345.

[5] Palczewski, K, Kumasaka, T, Hori, T, Behnke, C. A, Motoshima, H, Fox, B. A, Trong, I. L, Teller, D. C, Okada, T, Stenkamp, R. E, Yamamoto, M, & Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* (**2000**) 289, 739–45.

[6] Baldwin, J. M. Structure and function of receptors coupled to G proteins. *Curr. Opin. Cell. Biol.* (**1994**) 6, 180–190.

[7] Nederkoorn, P. H, Timmerman, H, & Donne-Op den Kelder, G. M. Does the ternary complex act as a secondary proton pump and a GTP synthase? *Trends Pharmacol. Sci.* (**1995**) 16, 156–161.

[8] Oliveira, L, Paiva, A. C, & Vriend, G. A common motif in G protein-coupled seven transmembrane helix receptors. *J. Comput. Aided Mol. Des.* (**1993**) 7, 649–658.

[9] van Rhee, A. M & Jacobson, K. A. Molecular Architecture of G protein-Coupled Receptors. *Drug Dev. Res.* (**1996**) 37, 1–38.

[10] Selbie, L. A & Hill, S. J. G protein-coupled-receptor cross-talk: the fine-tuning of multiple receptor-signalling pathways. *Trends Pharmacol. Sci.* (**1998**) 19, 87–93.

[11] Birnbaumer, L & Birnbaumer, M. Signal transduction by G proteins: 1994 edition. *J. Recept. Signal. Transduct. Res.* (**1995**) 15, 213–52.

[12] Robinson, W. E & Hagins, W. A. GTP hydrolysis in intact rod outer segments and the transmitter cycle in visual excitation. *Nature* (**1979**) 280, 398–400.

[13] Fung, B. K & Stryer, L. Surface density determination in membranes by fluorescence energy transfer. *Biochemistry* (**1978**) 17, 5241–8.

[14] Hohenegger, M, Mitterauer, T, Voss, T, Nanoff, C, & Freissmuth, M. Thiophosphorylation of the G protein beta subunit in human platelet. *Mol. Pharmacol.* (**1996**) 49, 73–80.

[15] Nederkoorn, P. H, Timmerman, H, Donne-Op den Kelder, G. M, Timms, D, Wilkinson, A. J, Kelly, D. R, Broadley, K. J, & Davies, R. H. GTP synthases. Proton pumping and phosphorylation in ligand-receptor-G alpha-protein complexes. *Receptors Channels* (**1996**) 4, 111–28.

[16] Topiol, S. The deletion model for the origin of receptors. *Trends Biochem. Sci.* (**1987**) 12, 419–421.

[17] Boyer, P. D. ATP synthase–past and future. *Biochim. Biophys. Acta.* (**1998**) 1365, 3–9.

[18] Boyer, P. D. Catalytic site forms and controls in ATP synthase catalysis. *Biochim. Biophys. Acta.* (**2000**) 1458, 252–62.

[19] Liri, T, Herzmark, P, Nakamoto, J. M, van Dop, C, & Bourne, H. R. Rapid GDP release from G-s-alpha in patients with gain and loss of endocrine function. *Nature* (**1994**) 371, 164–168.

[20] Nederkoorn, P. H, Vernooijs, P, Donne-Op den Kelder, G. M, Baerends, E. J, & Timmerman, H. A new model for the agonistic binding site on the histamine H2-receptor: the catalytic triad in serine proteases as a model for the binding site of histamine H2-receptor agonists. *J. Mol. Graph.* (**1994**) 12, 242–56.

[21] Lefkowitz, R. J, Cotecchia, S, Samama, P, & Costa, T. Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol. Sci.* (**1993**) 14, 303–7.

[22] Leurs, R, Smit, M. J, Meeder, R, ter Laak, A. M, & Timmerman, H. Lysine200 located in the fifth transmembrane domain of the histamine H1 receptor interacts with histamine but not with all H1 agonists. *Biochem. Biophys. Res. Commun.* (**1995**) 214, 110–7.

[23] ter Laak, A. M, Timmerman, H, Leurs, R, Nederkoorn, P. H, Smit, M. J, & Donne-Op den Kelder, G. M. Modelling and mutation studies on the histamine H1-receptor agonist binding site reveal different binding modes for H1- agonists: Asp116 (TM3) has a constitutive role in receptor stimulation. *J. Comput. Aided Mol. Des.* (**1995**) 9, 319–330.

[24] Turin, L. A spectroscopic mechanism for primary olfactory reception. *Chem. Senses.* (**1996**) 21, 773–91.

[25] Oliveira, L, Paiva, A. C, Sander, C, & Vriend, G. A common step for signal transduction in G protein-coupled receptors. *Trends Pharmacol. Sci.* (**1994**) 15, 170–172.

[26] Flower, D. R. Modelling G protein-coupled receptors for drug design. *Biochim. Biophys. Acta.* (**1999**) 1422, 207–34.

[27] Gudermann, T, Nurnberg, B, & Schultz, G. Receptors and G proteins as primary components of transmembrane signal transduction. Part 1. G protein-coupled receptors: structure and function. *J. Mol. Med.* (**1995**) 73, 51–63.

[28] Schwartz, T. W, Gether, U, Schambye, H. T, & Hjorth, S. A. Molecular mechanism of action of non-peptide ligands for peptide receptors. *Curr. Pharm. Design.* (**1995**) 1, 325–342.

[29] de Tullio, P, Delarge, J, & Pirotte, B. Therapeutic and chemical developments of cholecystokinin receptor ligands. *Exp. Opin. Invest. Drugs* (**2000**) 9, 129–146.

[30] Dolle, R. E. Comprehensive survey of chemical libraries yielding enzyme inhibitors, receptor agonists and antagonists, and other biologically active agents: 1992 through 1997. *Mol. Divers.* (**1998**) 3, 199–233.

[31] Beck-Sickinger, A. G. Structural Characterization and binding sites of G protein-coupled receptors. *Drug Design Today* (**1996**) 1, 502–513.

[32] Bikker, J. A, Trumpp-Kallmeyer, S, & Humblet, C. G protein coupled receptors: models, mutagenesis, and drug design. *J. Med. Chem.* (**1998**) 41, 2911–2927.

[33] Parmentier, M. L, Joly, C, Restituito, S, Bockaert, J, Grau, Y, & Pin, J. P. The G protein-coupling profile of metabotropic glutamate receptors, as determined with exogenous G proteins, is independent of their ligand recognition domain. *Mol. Pharmacol.* (**1998**) 53, 778–86.

[34] van Wagenen, B. C, Artmann, L. D, Hammerland, L. G, Hung, B. C. P, Johnston, J, Krapcho, K, Levinthal, C, Logan, M. A, Moe, S. T, Mueller, A. L, Simin, R, Smith, D. L, Storjohann, L, & Stormann, T. M. *Society for Neuroscience.* (**1998**) 24, 576.

[35] Elling, C. E & Schwartz, T. W. Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.* (**1996**) 15, 6213–9.

[36] Ji, H, Zheng, W, Zhang, Y, Catt, K. J, & Sandberg, K. Genetic transfer of a nonpeptide antagonist binding site to a previously unresponsive angiotensin receptor. *Proc. Natl. Acad. Sci. U. S. A.* (**1995**) 92, 9240–4.

[37] Prendergast, K, Adams, K, Greenlee, W. J, Nachbar, R. B, Patchett, A. A, & Underwood, D. J. Derivation of a 3D pharmacophore model for the angiotensin-II site one receptor. *J. Comput. Aided Mol. Des.* (**1994**) 8, 491–512.

[38] Lloyd, E. J & Andrews, P. R. A common structural model for central nervous system drugs and their receptors. *J. Med. Chem.* (**1986**) 29, 453–62.

[39] Tang, L, Ebrey, T. G, & Subramaniam, S. Sequences and Strcutures of Retinal Proteins. *Israel. J. Chem.* (**1995**) 35, 193–209.

[40] Sakmar, T. P. Rhodopsin: a prototypical G protein-coupled receptor. *Prog. Nucleic. Acid. Res. Mol. Biol.* (**1998**) 59, 1–34.

[41] Henderson, R & Unwin, P. N. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* (**1975**) 257, 28–32.

[42] Pebay-Peyroula, E, Rummel, G, Rosenbusch, J. P, & Landau, E. M. X-ray structure of bacteriorhodopsin at 2. 5 angstroms from microcrystals grown in lipidic cubic phases. *Science* (**1997**) 277, 1676–1681.

[43] Gouaux, E. It's not just a phase: crystallization and X-ray structure determination of bacteriorhodopsin in lipidic cubic phases. *Structure.* (**1998**) 6, 5–10.

[44] Moody, M. F. (**1990**) in *Biophysical. Electron. Microscopy.*, eds. Hawkes, P. W & Valdre, U. (Academic Press, London).

[45] Misell, D. L & Brown, E. B. (**1987**) *Electron Diffraction: an Introduction for Biologist*, Practical Methods in Electron Microscopy. (Elsevier, Amsterdam) Vol. 12.

[46] Unwin, P. N & Henderson, R. Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* (**1975**) 94, 425–440.

[47] Amos, L. A, Henderson, R, & Unwin, P. N. T. Three-dimensional structure determination by electron microscopy of two dimensional crystals. *Proc. Biophys. Molec. Biol.* (**1982**) 39, 183–231.

[48] Henderson, R, Baldwin, J. M, Ceska, T. A, Zemlin, F, Beckmann, E, & Downing, K. H. Model for the structure of bacteriorhodopsin based on high- resolution electron cryo-microscopy. *J. Mol. Biol.* (**1990**) 213, 899–929.

[49] Grigorieff, N, Ceska, T. A, Downing, K. H, Baldwin, J. M, & Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* (**1996**) 259, 393–421.

[50] Kimura, Y, Vassylyev, D. G, Miyazawa, A, Kidera, A, Matsushima, M, Mitsuoka, K, Murata, K, Hirai, T, & Fujiyoshi, Y. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* (**1997**) 389, 206–211.

[51] Havelka, W. A, Henderson, R, Heymann, J. A, & Oesterhelt, D. Projection structure of halorhodopsin from Halobacterium halobium at 6 A resolution obtained by electron cryo-microscopy. *J. Mol. Biol.* (**1993**) 234, 837–846.

[52] Havelka, W. A, Henderson, R, & Oesterhelt, D. Three-dimensional structure of halorhodopsin at 7 A resolution. *J. Mol. Biol.* (**1995**) 247, 726–738.

[53] Landau, E. M & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* (**1996**) 93, 14532–5.

[54] Chiu, M. L, Nollert, P, Loewen, M. C, Belrhali, H, Pebay-Peyroula, E, Rosenbusch, J. P, & Landau, E. M. Crystallization in cubo: general applicability to membrane proteins. *Acta. Crystallogr. D. Biol. Crystallogr.* (**2000**) 56, 781–4.

[55] Rummel, G, Hardmeyer, A, Widmer, C, Chiu, M. L, Nollert, P, Locher, K. P, Pedruzzi, I, Landau, E. M, & Rosenbusch, J. P. Lipidic Cubic Phases: New Matrices for the Three-Dimensional Crystallization of Membrane Proteins. *J. Struct. Biol.* (**1998**) 121, 82–91.

[56] Luecke, H, Richter, H. T, & Lanyi, J. K. Proton transfer pathways in bacteriorhodopsin at 2. 3 angstrom resolution. *Science* (**1998**) 280, 1934–7.

[57] Schertler, G. F, Villa, C, & Henderson, R. Projection structure of rhodopsin. *Nature* (**1993**) 362, 770–772.

[58] Schertler, G. F & Hargrave, P. A. Projection structure of frog rhodopsin in two crystal forms. *Proc. Natl. Acad. Sci. U. S. A.* (**1995**) 92, 11578–11582.

[59] Davies, A, Schertler, G. F, Gowen, B. E, & Saibil, H. R. Projection structure of an invertebrate rhodopsin. *J. Struct. Biol.* (**1996**) 117, 36–44.

[60] Baldwin, J. M. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* (**1993**) 12, 1693–1703.

[61] Bourne, H. R & Meng, E. C. Structure. Rhodopsin sees the light. *Science* (**2000**) 289, 733–4.

[62] Unger, V. M, Hargrave, P. A, Baldwin, J. M, & Schertler, G. F. Arrangement of rhodopsin transmembrane alpha-helices. *Nature* (**1997**) 389, 203–206.

[63] Unger, V. M & Schertler, G. F. Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. *Biophys. J.* (**1995**) 68, 1776–1786.

[64] Krebs, A, Villa, C, Edwards, P. C, & Schertler, G. F. X. Characterisation of an improved two-dimensional p22121 crystal from bovine rhodopsin. *J. Mol. Biol.* (**1998**) 282, 991–1003.

[65] Schertler, G. F. Structure of rhodopsin. *Eye.* (**1998**) 12, 504–10.

[66] Baldwin, J. M, Schertler, G. F, & Unger, V. M. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G protein-coupled receptors. *J. Mol. Biol.* (**1997**) 272, 144–164.

[67] Donnelly, D & Cogdell, R. J. Predicting the point at which transmembrane helices protrude from the bilayer: a model of the antenna complexes from photosynthetic bacteria. *Protein Eng.* (**1993**) 6, 629–635.

[68] Reithmeier, R. A. Characterization and modeling of membrane proteins using sequence analysis. *Curr. Opin. Struct. Biol.* (**1995**) 5, 491–500.

[69] Noble, F & Roques, B. P. CCK-B receptor: chemistry, molecular biology, biochemistry and pharmacology. *Prog. Neurobiol.* (**1999**) 58, 349–79.

[70] Wank, S. A. G protein-coupled receptors in gastrointestinal physiology. I. CCK receptors: an exemplary family. *Am. J. Physiol.* (**1998**) 274, G607–13.

[71] Silvente-Poirot, S, Dufresne, M, Vaysse, N, & Fourmy, D. The peripheral cholecystokinin receptors. *Eur. J. Biochem.* (**1993**) 215, 513–29.

[72] van Gunsteren, W. F & Berendsen, H. J. C. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.* (**1990**) 29, 992–1023.

[73] Karplus, M & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* (**1990**) 347, 631–9.

[74] Guba, W, Haessner, R, Breipohl, G, Henke, S, Knolle, J, Santagada, V, & Kessler, H. Combined approach of NMR and molecular dynamics within a biphasic membrane mimetic. conformation and orientation of the bradykinin antagonist Hoe 140. *J. Am. Chem. Soc.* (**1994**) 116, 7532–7540.

[75] Guba, W & Kessler, H. A novel computational mimetic of biological membranes in molecular dynamics simulations. *J. Phys. Chem.* (**1994**) 98, 23–27.

[76] Dayhoff, M. O, Schwartz, R. M, & Orcutt, B. C. (**1978**) in *Atlas of Protein Sequence and Structure*. (Natl. Biomed. Res. Found., Washington, DC) Vol. 5 suppl. 3, pp. 345–352.

[77] Schwartz, R. M & Dayhoff, M. O. (**1978**) in *Atlas of Protein Sequence and Structure*. (Natl. Biomed. Res. Found., Washington, DC) Vol. 5 suppl. 3, pp. 353–358.

[78] Kyte, J & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* (**1982**) 157, 105–132.

[79] Ballesteros, J. A & Weinstein, H. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G protein-Coupled Receptors. *Meth. Neurosc.* (**1995**) 25, 366–428.

[80] Crimi, M & Esposti, M. D. Structural predictions for membrane proteins: the dilemma of hydrophobicity scales. *Trends Biochem. Sci.* (**1991**) 16, 119.

[81] Sipos, L & von Heijne, G. Predicting the topology of eukaryotic membrane proteins. *Eur. J. Biochem.* (**1993**) 213, 1333–1340.

[82] Rose, G. D & Dworkin, J. E. (**1989**) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum Press, New York), pp. 625–633.

[83] Esposti, M. D, Crimi, M, & Venturoli, G. A critical evaluation of the hydropathy profile of membrane proteins. *Eur. J. Biochem.* (**1990**) 190, 207–219.

[84] Eisenberg, D & McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* (**1986**) 319, 199–203.

[85] Zhang, D & Weinstein, H. Polarity conserved positions in transmembrane domains of G- protein coupled receptors and bacteriorhodopsin. *FEBS Lett.* (**1994**) 337, 207–212.

[86] Wolfenden, R, Andersson, L, Cullis, P. M, & Southgate, C. C. Affinities of amino acid side chains for solvent water. *Biochemistry* (**1981**) 20, 849–855.

[87] Eisenberg, D, Weiss, R. M, & Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. U. S. A.* (**1984**) 81, 140–144.

[88] Hopp, T. P & Woods, K. R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. U. S. A.* (**1981**) 78, 3824–3828.

[89] Abraham, D. J & Leo, A. J. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins* (**1987**) 2, 130–52.

[90] Roseman, M. A. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* (**1988**) 200, 513–22.

[91] Sweet, R. M & Eisenberg, D. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. Mol. Biol.* (**1983**) 171, 479–88.

[92] Zimmerman, J. M, Eliezer, N, & Simha, R. The Characterization of Amino Acid Sequences in Proteins by Statistical Methods. *J. Theoret. Biol.* (**1968**) 21, 170–201.

[93] Janin, J. Surface and inside volumes in globular proteins. *Nature* (**1979**) 277, 491–2.

[94] Chou, P. Y & Fasman, G. D. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* (**1978**) 47, 251–276.

[95] von Heijne, G. The distribution of positvely charged residues in bacterial inner membrane proteins correlates with the transmembrane topology. *EMBO J.* (**1986**) 5, 3021–3027.

[96] Williams, K. A & Deber, C. M. Proline residues in transmembrane helices: structural or dynamic role? *Biochemistry* (**1991**) 30, 8919–8923.

[97] von Heijne, G. Proline kinks in transmembrane alpha-helices. *J. Mol. Biol.* (**1991**) 218, 499–503.

[98] Landolt-Marticorena, C, Williams, K. A, Deber, C. M, & Reithmeier, R. A. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J. Mol. Biol.* (**1993**) 229, 602–608.

[99] Schiffer, M & Edmundson, A. B. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* (**1967**) 7, 121–135.

[100] Eisenberg, D, Weiss, R. M, & Terwilliger, T. C. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* (**1982**) 299, 371–374.

[101] Eisenberg, D, Wesson, M, & Wilcox, W. (**1989**) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum Press, New York), pp. 635–646.

[102] Dunnill, P. The use of helical net-diagrams to represent protein structures. *Biophys. J.* (**1968**) 8, 865–875.

[103] Finer-Moore, J, Bazan, J. F, Rubin, J, & Stroud, R. M. (**1989**) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum Press, New York), pp. 719–759.

[104] von Heijne, G & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* (**1988**) 174, 671–678.

[105] Rost, B, Sander, C, & Schneider, R. PHD - an automatic mail server for protein secondary structure prediction. *CABIOS* (**1994**) 10, 53–60.

[106] Rost, B, Casadio, R, Fariselli, P, & Sander, C. Prediction of helical transmembrane segments at 95% accuracy. *Protein Science* (**1995**) 4, 521–533.

[107] Donnelly, D, Overington, J. P, Ruffle, S. V, Nugent, J. H, & Blundell, T. L. Modeling alpha-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. *Protein Sci.* (**1993**) 2, 55–70.

[108] Donnelly, D, Findlay, J. B, & Blundell, T. L. The evolution and structure of aminergic G protein-coupled receptors. *Receptors. Channels.* (**1994**) 2, 61–78.

[109] von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* (**1992**) 225, 487–494.

[110] Hecht, M. H, Richardson, J. S, Richardson, D. C, & Ogden, R. C. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* (**1990**) 249, 884–891.

[111] Kamtekar, S, Schiffer, J. M, Xiong, H, Babik, J. M, & Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* (**1993**) 262, 1680–1685.

[112] Kolbe, M, Besir, H, Essen, L. O, & Oesterhelt, D. Structure of the light-driven chloride pump halorhodopsin at 1. 8 A. *Science* (**2000**) 288, 1390–6.

[113] Woolf, T. B & Roux, B. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proc. Natl. Acad. Sci. U. S. A.* (**1994**) 91, 11631–11635.

[114] Birnbaumer, L, Abramowitz, J, & Brown, A. M. Receptor-effector coupling by G proteins. *Biochim. Biophys. Acta.* (**1990**) 1031, 163–224.

[115] Chang, R. S, Lotti, V. J, Monaghan, R. L, Birnbaum, J, Stapley, E. O, Goetz, M. A, Albers-Schonberg, G, Patchett, A. A, Liesch, J. M, Hensens, O. D, & et al. A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from Aspergillus alliaceus. *Science* (**1985**) 230, 177–9.

[116] Hoflack, J, Trumpp-Kallmeyer, S, & Hibert, M. Re-evaluation of bacteriorhodopsin as a model for G protein- coupled receptors. *Trends Pharmacol. Sci.* (**1994**) 15, 7–9.

[117] IJzerman, A. P, Galen, P. J. V, & Jacobson, K. A. Molecular modeling of adenosine receptors. I. The ligand binding site on the A1 receptor. *Drug Des. Discov.* (**1992**) 9, 49–67.

[118] Trumpp-Kallmeyer, S, Hoflack, J, Bruinvels, A, & Hibert, M. Modeling of G protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. *J. Med. Chem.* (**1992**) 35, 3448–3462.

[119] Cronet, P, Sander, C, & Vriend, G. Modeling of transmembrane seven helix bundles. *Protein Eng.* (**1993**) 6, 59–64.

[120] Lutz, J, Romano-Gotsch, R, Escrieut, C, Fourmy, D, Matha, B, Müller, G, Kessler, H, & Moroder, L. Mapping of ligand binding sites of the cholecystokinin-B/gastrin receptor with lipo-gastrin peptides and molecular modeling. *Biopolymers* (**1997**) 41, 799–817.

[121] MaloneyHuss, K & Lybrand, T. P. Three-dimensional structure for the beta 2 adrenergic receptor protein based on computer modeling studies. *J. Mol. Biol.* (**1992**) 225, 859–871.

[122] Herzyk, P & Hubbard, R. E. Automated method for modeling seven-helix transmembrane receptors from experimental data. *Biophys. J.* (**1995**) 69, 2419–2442.

[123] van Neuren, A. S, Müller, G, Klebe, G, & Moroder, L. Molecular modelling studies on G protein-coupled receptors: from sequence to structure? *J. Recept. Signal. Transduct. Res.* (**1999**) 19, 341–53.

[124] Nakata, H, Matsui, T, Ito, M, Taniguchi, T, Naribayashi, Y, Arima, N, Nakamura, A, Kinoshita, Y, Chihara, K, Hosoda, S, & et al. Cloning and characterization of gastrin receptor from ECL carcinoid tumor of Mastomys natalensis. *Biochem. Biophys. Res. Commun.* (**1992**) 187, 1151–7.

[125] Wank, S. A, Pisegna, J. R, & de Weerth, A. Brain and gastrointestinal cholecystokinin receptor family: structure and functional expression. *Proc. Natl. Acad. Sci. U. S. A.* (**1992**) 89, 8691–5.

[126] Lee, Y. M, Beinborn, M, McBride, E. W, Lu, M, jr, L. F. K, & Kopin, A. S. The human brain cholecystokinin-B/gastrin receptor. Cloning and characterization. *J. Biol. Chem.* (**1993**) 268, 8164–9.

[127] Blandizzi, C, Song, I, & Yamada, T. Molecular cloning and structural analysis of the rabbit gastrin/CCK-B receptor gene. *Biochem. Biophys. Res. Commun.* (**1994**) 202, 947–53.

[128] Kopin, A. S, Lee, Y. M, McBride, E. W, Miller, L. J, Lu, M, Lin, H. Y, Kolakowski, L. F, & Beinborn, M. Expression cloning and characterization of the canine parietal cell gastrin receptor. *Proc. Natl. Acad. Sci. U. S. A.* (**1992**) 89, 3605–9.

[129] Weerth, A. D, Pisegna, J. R, & Wank, S. A. Guinea pig gallbladder and pancreas possess identical CCK-A receptor subtypes: receptor cloning and expression. *Am. J. Physiol.* (**1993**) 265, G1116–21.

[130] Ulrich, C. D, Ferber, I, Holicky, E, Hadac, E, Buell, G, & Miller, L. J. Molecular cloning and functional expression of the human gallbladder cholecystokinin A receptor. *Biochem. Biophys. Res. Commun.* (**1993**) 193, 204–11.

[131] Wank, S. A, Harkins, R, Jensen, R. T, Shapira, H, de Weerth, A, & Slattery, T. Purification, molecular cloning, and functional expression of the cholecystokinin receptor from rat pancreas. *Proc. Natl. Acad. Sci. U. S. A.* (**1992**) 89, 3125–9.

[132] Schmitz, F, Pratt, D. S, Wu, M. J, Kolakowski, L. F, Beinborn, M, & Kopin, A. S. Identification of cholecystokinin-B/gastrin receptor domains that confer high gastrin affinity: utilization of a novel Xenopus laevis cholecystokinin receptor. *Mol. Pharmacol.* (**1996**) 50, 436–41.

[133] Killian, J. A, Timmermans, J. W, Keur, S, & de Kruijff, B. The tryptophans of gramicidin are essential for the lipid structure modulating effect of the peptide. *Biochim. Biophys. Acta.* (**1985**) 820, 154–156.

[134] Dunbrack, R. L & Karplus, M. Backbone-dependent rotamer library for proteins. Application to side- chain prediction. *J. Mol. Biol.* (**1993**) 230, 543–74.

[135] Dunbrack, R. L & Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* (**1994**) 1, 334–40.

[136] Dunbrack, R. L & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* (**1997**) 6, 1661–81.

[137] Bower, M. J, Cohen, F. E, & Dunbrack, R. L. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* (**1997**) 267, 1268–82.

[138] Guba, W. Modellierung der nativen Umgebung von Molekülen in Moleküldynamik-Simulationen. Solvens, Modellmembran, Zweiphasen-Membranmimetikum. *Thesis of the Institut für Organische Chemie und Biochemie, Technischen Universität München* (**1994**) pp. 1–171.

[139] Dixon, R. A, Sigal, I. S, Candelore, M. R, Register, R. B, Scattergood, W, Rands, E, & Strader, C. D. Structural features required for ligand binding to the beta-adrenergic receptor. *EMBO J.* (**1987**) 6, 3269–75.

[140] Leong, S. R, Kabakoff, R. C, & Hebert, C. A. Complete mutagenesis of the extracellular domain of interleukin-8 (IL- 8) type A receptor identifies charged residues mediating IL-8 binding and signal transduction. *J. Biol. Chem.* (**1994**) 269, 19343–8.

[141] Silvente-Poirot, S, Escrieut, C, & Wank, S. A. Role of the extracellular domains of the cholecystokinin receptor in agonist binding. *Mol. Pharmacol.* (**1998**) 54, 364–71.

[142] van Klompenburg, W & de Kruijff, B. The role of anionic lipids in protein insertion and translocation in bacterial membranes. *J. Membr. Biol.* (**1998**) 162, 1–7.

[143] van Klompenburg, W, Nilsson, I, von Heijne, G, & de Kruijff, B. Anionic phospholipids are determinants of membrane protein topology. *EMBO J.* (**1997**) 16, 4261–6.

[144] Killian, J. A & de Kruijff, B. Importance of hydration for gramicidin-induced hexagonal HII phase formation in dioleoylphosphatidylcholine model membranes. *Biochemistry* (**1985**) 24, 7890–7898.

[145] Tieleman, D. P, Forrest, L. R, Sansom, M. S. P, & Berendsen, H. J. C. Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: molecular dynamics simulations. *Biochemistry* (**1998**) 37, 17554–61.

[146] Koradi, R, Billeter, M, & Wthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics.* (**1996**) 14, 51–55.

[147] Kopin, A. S, McBride, E. W, Quinn, S. M, Kolakowski, L. F, & Beinborn, M. The role of the cholecystokinin-B/gastrin receptor transmembrane domains in determining affinity for subtype-selective ligands. *J. Biol. Chem.* (**1995**) 270, 5019–5023.

[148] Blaker, M, Ren, Y, Seshadri, L, McBride, E. W, Beinborn, M, & Kopin, A. S. CCK-B/Gastrin receptor transmembrane domain mutations selectively alter synthetic agonist efficacy without affecting the activity of endogenous peptides. *Mol. Pharmacol.* (**2000**) 58, 399–406.

[149] Blaker, M, Ren, Y, Gordon, M. C, Hsu, J. E, Beinborn, M, & Kopin, A. S. Mutations within the cholecystokinin-B/gastrin receptor ligand 'pocket' interconvert the functions of nonpeptide agonists and antagonists. *Mol. Pharmacol.* (**1998**) 54, 857–63.

[150] Jagerschmidt, A, Guillaume, N, Goudreau, N, Maigret, B, & Roques, B. P. Mutation of Asp100 in the second transmembrane domain of the cholecystokinin B receptor increases antagonist binding and reduces signal transduction. *Mol. Pharmacol.* (**1995**) 48, 783–9.

[151] Jagerschmidt, A, Guillaume-Rousselet, N, Vikland, M. L, Goudreau, N, Maigret, B, & Roques, B. P. His381 of the rat CCK-B receptor is essential for CCK-B versus CCK-A receptor antagonist selectivity. *Eur. J. Pharmacol.* (**1996**) 296, 97–106.

[152] Gales, C, Kowalski-Chauvel, A, Dufour, M. N, Seva, C, Moroder, L, Pradayrol, L, Vaysse, N, Fourmy, D, & Silvente-Poirot, S. Mutation of Asn-391 within the conserved NPXXY motif of the cholecystokinin B receptor abolishes Gq protein activation without affecting its association with the receptor. *J. Biol. Chem.* (**2000**) 275, 17321–7.

[153] Silvente-Poirot, S & Wank, S. A. A segment of five amino acids in the second extracellular loop of the cholecystokinin B receptor is essential for selectivity of the peptide agonist gastrin. *J. Biol. Chem.* (**1996**) 271, 14698–14706.

[154] Beinborn, M, Quinn, S. M, & Kopin, A. S. Minor modifications of a cholecystokinin-B/gastrin receptor non- peptide antagonist confer a broad spectrum of functional properties. *J. Biol. Chem.* (**1998**) 273, 14146–14151.

[155] Makovec, F & D'Amato, M. CCK-B/Gastrin receptor antagonists as potential drugs for peptic ulcer therapy. *Drug Design Today* (**1997**) 2, 283–293.

[156] Trivedi, B. K. Ligands for cholecystokinin receptors: recent developments. *Curr. Opin. Therap. Patents.* (**1994**) 4, 31–44.

[157] Trivedi, B. K. Cholecystokinin receptor antagonists: current status. *Curr. Med. Chem.* (**1994**) 1, 313–327.

[158] Revel, L & Makovec, F. Update on nonpeptide CCK-B receptor antagonists. *Drugs of the Future* (**1998**) 23, 751–766.

[159] Blommaert, A. G, Dhotel, H, Ducos, B, Durieux, C, Goudreau, N, Bado, A, Garbay, C, & Roques, B. P. Structure-based design of new constrained cyclic agonists of the cholecystokinin CCK-B receptor. *J. Med. Chem.* (**1997**) 40, 647–658.

[160] Tilley, J. W, Danho, W, Shiuey, S. J, Kulesha, I, Swistok, J, Makofske, R, Michalewsky, J, Triscari, J, Nelson, D, Weatherford, S, & et al. Analogs of Ac-CCK-7 incorporating dipeptide mimics in place of Met28-Gly29. *J. Med. Chem.* (**1992**) 35, 3774–83.

[161] Danho, W, Makofske, R. C, Swistok, J, Michalewsky, J, Gabriel, T, Marks, N, Berg, M. J, Baird, L, Geiler, V, Mackie, G, & et al. Synthesis and biological activity of novel, potent and long-acting analogs of AC-CCK-7 with high affinity for peripheral (type A) receptors. *Pept. Res.* (**1991**) 4, 59–65.

[162] Bernad, N, Burgaud, B. G, Horwell, D. C, Lewthwaite, R. A, Martinez, J, & Pritchard, M. C. The design and synthesis of the high efficacy, non-peptide CCK1 receptor agonist PD170292. *Bioorg. Med. Chem. Lett.* (**2000**) 10, 1245–1248.

[163] Bignon, E, Alonso, R, Arnone, M, Boigegrain, R, Brodin, R, Gueudet, C, Heaulme, M, Keane, P, Landi, M, Molimard, J. C, Olliero, D, Poncelet, M, Seban, E, Simiand, J, Soubrie, P, Pascal, M, Maffrand, J. P, & Fur, G. L. SR146131: a new potent, orally active, and selective nonpeptide cholecystokinin subtype 1 receptor agonist. II. In vivo pharmacological characterization. *J. Pharmacol. Exp. Therap.* (**1999**) 289, 752–61.

[164] Hirst, G. C, Aquino, C, Birkemo, L, Croom, D. K, Dezube, M, Dougherty, R. W, Ervin, G. N, Grizzle, M. K, Henke, B, James, M. K, Johnson, M. F, Momtahen, T, Queen, K. L, Sherrill, R. G, Szewczyk, J, Willson, T. M, & Sugg, E. E. Discovery of 1,5-benzodiazepines with peripheral cholecystokinin (CCK-A) receptor agonist activity (II): Optimization of the C3 amino substituent. *J. Med. Chem.* (**1996**) 39, 5236–45.

[165] Henke, B. R, Willson, T. M, Sugg, E. E, Croom, D. K, Dougherty, R. W, Queen, K. L, Birkemo, L. S, Ervin, G. N, Grizzle, M. K, Johnson, M. F, & James, M. K. 3-(1H-

indazol-3-ylmethyl)-1,5-benzodiazepines: CCK-A agonists that demonstrate oral activity as satiety agents. *J. Med. Chem.* (**1996**) 39, 2655–8.

[166] Pierson, M. E, Comstock, J. M, Simmons, R. D, Kaiser, F, Julien, R, Zongrone, J, & Rosamond, J. D. Synthesis and biological evaluation of potent, selective, hexapeptide CCK-A agonist anorectic agents. *J. Med. Chem.* (**1997**) 40, 4302–7.

[167] Pierson, M. E, Comstock, J. M, Simmons, R. D, Julien, R, Kaiser, F, & Rosamond, J. D. CCK peptides with combined features of hexa- and tetrapeptide CCK-A agonists. *J. Med. Chem.* (**2000**) 43, 2350–5.

[168] Bignon, E, Bachy, A, Boigegrain, R, Brodin, R, Cottineau, M, Gully, D, Herbert, J. M, Keane, P, Labie, C, Molimard, J. C, Olliero, D, Oury-Donat, F, Petereau, C, Prabonnaud, V, Rockstroh, M. P, Schaeffer, P, Servant, O, Thurneyssen, O, Soubrie, P, Pascal, M, Maffrand, J. P, & Fur, G. L. SR146131: a new potent, orally active, and selective nonpeptide cholecystokinin subtype 1 receptor agonist. I. In vitro studies. *J. Pharmacol. Exp. Therap.* (**1999**) 289, 742–51.

[169] Aquino, C. J, Armour, D. R, Berman, J. M, Birkemo, L. S, Carr, R. A, Croom, D. K, Dezube, M, Dougherty, R. W, Ervin, G. N, Grizzle, M. K, Head, J. E, Hirst, G. C, James, M. K, Johnson, M. F, Miller, L. J, Queen, K. L, Rimele, T. J, Smith, D. N, & Sugg, E. E. Discovery of 1,5-benzodiazepines with peripheral cholecystokinin (CCK-A) receptor agonist activity. 1. Optimization of the agonist "trigger". *J. Med. Chem.* (**1996**) 39, 562–9.

[170] Willson, T. M, Henke, B. R, Momtahen, T. M, Myers, P. L, Sugg, E. E, Unwalla, R. J, Croom, D. K, Dougherty, R. W, Grizzle, M. K, Johnson, M. F, Queen, K. L, Rimele, T. J, Yingling, J. D, & James, M. K. 3-[2-(N-phenylacetamide)]-1,5-benzodiazepines: orally active, binding selective CCK-A agonists. *J. Med. Chem.* (**1996**) 39, 3030–4.

[171] Henke, B. R, Aquino, C. J, Birkemo, L. S, Croom, D. K, Dougherty, R. W, Ervin, G. N, Grizzle, M. K, Hirst, G. C, James, M. K, Johnson, M. F, Queen, K. L, Sherrill, R. G, Sugg, E. E, Suh, E. M, Szewczyk, J. W, Unwalla, R. J, Yingling, J, & Willson, T. M. Optimization of 3-(1H-indazol-3-ylmethyl)-1,5-benzodiazepines as potent, orally active CCK-A agonists. *J. Med. Chem.* (**1997**) 40, 2706–25.

[172] Tilley, J. W, Danho, W, Lovey, K, Wagner, R, Swistok, J, Makofske, R, Michalewsky, J, Triscari, J, Nelson, D, & Weatherford, S. Carboxylic acids and tetrazoles as isosteric replacements for sulfate in cholecystokinin analogues. *J. Med. Chem.* (**1991**) 34, 1125–36.

[173] Goudreau, N, Weng, J. H, & Roques, B. P. Conformational analysis of CCK-B agonists using 1H-NMR and restrained molecular dynamics: comparison of biologically active Boc-Trp-(N-Me) Nle-Asp-Phe-NH2 and inactive Boc-Trp-(N-Me)Phe-Asp-Phe-NH2. *Biopolymers* (**1994**) 34, 155–69.

[174] Lignon, M. F, Galas, M. C, Rodriguez, M, Laur, J, Aumelas, A, & Martinez, J. A synthetic peptide derivative that is a cholecystokinin receptor antagonist. *J. Biol. Chem.* (**1987**) 262, 7226–31.

[175] Hahne, W. F, Jensen, R. T, Lemp, G. F, & Gardner, J. D. Proglumide and benzotript: members of a different class of cholecystokinin receptor antagonists. *Proc. Natl. Acad. Sci. U. S. A.* (**1981**) 78, 6304–8.

[176] Makovec, F, Bani, M, Cereda, R, Chiste, R, Pacini, M. A, Revel, L, Rovati, L. A, Rovati, L. C, & Setnikar, I. Pharmacological properties of lorglumide as a member of a new class of cholecystokinin antagonists. *Arzneimittelforschung.* (**1987**) 37, 1265–8.

[177] Setnikar, I, Bani, M, Cereda, R, Chiste, R, Makovec, F, Pacini, M. A, Revel, L, Rovati, L. C, & Rovati, L. A. Pharmacological characterisation of a new potent and specific nonpolypeptidic cholecystokinin antagonist. *Arzneimittelforschung.* (**1987**) 37, 703–7.

[178] van der Bent, A, Blommaert, A. G, Melman, C. T, IJzerman, A. P, van Wijngaarden, I, & Soudijn, W. Hybrid cholecystokinin A antagonists based on molecular modeling of lorglumide and L-364,718. *J. Med. Chem.* (**1992**) 35, 1042–1049.

[179] Boden, P. R, Higginbottom, M, Hill, D. R, Horwell, D. C, Hughes, J, Rees, D. C, Roberts, E, Singh, L, Suman-Chauhan, N, & Woodruff, G. N. Cholecystokinin dipeptoid antagonists: design, synthesis, and anxiolytic profile of some novel CCK-A and CCK-B selective and "mixed" CCK-A/CCK-B antagonists. *J. Med. Chem.* (**1993**) 36, 552–65.

[180] Martin-Martinez, M, Bartolome-Nebreda, J. M, Gomez-Monterrey, I, Gonzalez-Muniz, R, Garcia-Lopez, M. T, Ballaz, S, Barber, A, Fortuno, A, Rio, J. D, & Herranz, R. Synthesis and stereochemical structure-activity relationships of 1,3-dioxoperhydropyrido[1,2-c]pyrimidine derivatives: potent and selective cholecystokinin A receptor antagonists. *J. Med. Chem.* (**1997**) 40, 3402–7.

[181] Makovec, F, Chiste, R, Bani, M, Pacini, M. A, Setnikar, I, & Rovati, L. A. New glutaramic acid derivatives with potent competitive and specific cholecystokinin-antagonistic activity. *Arzneimittelforschung.* (**1985**) 35, 1048–51.

[182] Evans, B. E, Rittle, K. E, Bock, M. G, DiPardo, R. M, Freidinger, R. M, Whitter, W. L, Lundell, G. F, Veber, D. F, Anderson, P. S, Chang, R. S, & et al. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* (**1988**) 31, 2235–46.

[183] Bock, M. G, DiPardo, R. M, Evans, B. E, Rittle, K. E, Whitter, W. L, Garsky, V. M, Gilbert, K. F, Leighton, J. L, Carson, K. L, Mellin, E. C, & et al. Development of 1,4-benzodiazepine cholecystokinin type B antagonists. *J. Med. Chem.* (**1993**) 36, 4276–92.

[184] Bock, M. G, DiPardo, R. M, Evans, B. E, Rittle, K. E, Whitter, W. L, Veber, D. E, Anderson, P. S, & Freidinger, R. M. Benzodiazepine gastrin and brain cholecystokinin receptor ligands: L-365,260. *J. Med. Chem.* (**1989**) 32, 13–6.

[185] Bock, M. G, DiPardo, R. M, Evans, B. E, Rittle, K. E, Veber, D. F, Freidinger, R. M, Chang, R. S, & Lotti, V. J. Cholecystokinin antagonists. Synthesis and biological evaluation of 4-substituted 4H-[1,2,4]triazolo[4,3-a][1,4]benzodiazepines. *J. Med. Chem.* (**1988**) 31, 176–81.

[186] Bock, M. G, DiPardo, R. M, Evans, B. E, Rittle, K. E, Freidinger, R. M, Chang, R. S, & Lotti, V. J. Cholecystokinin antagonists. Synthesis and biological evaluation of 3-substituted 1,4-benzodiazepin-2-amines. *J. Med. Chem.* (**1988**) 31, 264–8.

[187] Evans, B. E, Rittle, K. E, Bock, M. G, DiPardo, R. M, Freidinger, R. M, Whitter, W. L, Gould, N. P, Lundell, G. F, Homnick, C. F, Veber, D. F, & et al. Design of nonpeptidal ligands for a peptide receptor: cholecystokinin antagonists. *J. Med. Chem.* (**1987**) 30, 1229–39.

[188] Evans, B. E, Bock, M. G, Rittle, K. E, DiPardo, R. M, Whitter, W. L, Veber, D. F, Anderson, P. S, & Freidinger, R. M. Design of potent, orally effective, nonpeptidal antagonists of the peptide hormone cholecystokinin. *Proc. Natl. Acad. Sci. U. S. A.* (**1986**) 83, 4918–22.

[189] Makovec, F, Peris, W, Revel, L, Giovanetti, R, Mennuni, L, & Rovati, L. C. Structure-antigastrin activity relationships of new (R)-4- benzamido-5- oxopentanoic acid derivatives. *J. Med. Chem.* (**1992**) 35, 28–38.

[190] Makovec, F, Peris, W, Frigerio, S, Giovanetti, R, Letari, O, Mennuni, L, & Revel, L. Structure-antigastrin activity relationships of new spiroglumide amido acid derivatives. *J. Med. Chem.* (**1996**) 39, 135–142.

[191] Horwell, D. C, Hughes, J, Hunter, J. C, Pritchard, M. C, Richardson, R. S, Roberts, E, & Woodruff, G. N. Rationally designed "dipeptoid" analogues of CCK. alpha-Methyltryptophan derivatives as highly selective and orally active gastrin and CCK-B antagonists with potent anxiolytic properties. *J. Med. Chem.* (**1991**) 34, 404–14.

[192] Trivedi, B. K, Padia, J. K, Holmes, A, Rose, S, Wright, D. S, Hinton, J. P, Pritchard, M. C, Eden, J. M, Kneen, C, Webdale, L, Suman-Chauhan, N, Boden, P, Singh, L, Field, M. J, & Hill, D. Second generation "peptoid" CCK-B receptor antagonists: identification and development of N-(adamantyloxycarbonyl)-alpha- methyl-(R)-tryptophan derivative (CI-1015) with an improved pharmacokinetic profile. *J. Med. Chem.* (**1998**) 41, 38–45.

[193] Corringer, P. J, Weng, J. H, Ducos, B, Durieux, C, Boudeau, P, Bohme, A, & Roques, B. P. CCK-B agonist or antagonist activities of structurally hindered and peptidase-resistant Boc-CCK4 derivatives. *J. Med. Chem.* (**1993**) 36, 166–72.

[194] Blommaert, A. G, Weng, J. H, Dorville, A, McCort, I, Ducos, B, Durieux, C, & Roques, B. P. Cholecystokinin peptidomimetics as selective CCK-B antagonists: design, synthesis, and in vitro and in vivo biochemical properties. *J. Med. Chem.* (**1993**) 36, 2868–77.

[195] Weng, J. H, Blommaert, A. G, Moizo, L, Bado, A, Ducos, B, Bohme, A, Garbay, C, & Roques, B. P. Role of N- and C-terminal substituents on the CCK-B agonist-antagonist pharmacological profile of Boc-Trp-Phg-Asp-Nal-NH2 derivatives. *Bioorg. Med. Chem.* (**1996**) 4, 563–73.

[196] Bellier, B, McCort-Tranchepain, I, Ducos, B, Danascimento, S, Meudal, H, Noble, F, Garbay, C, & Roques, B. P. Synthesis and biological properties of new constrained CCK-B antagonists: discrimination of two affinity states of the CCK-B receptor on transfected CHO cells. *J. Med. Chem.* (**1997**) 40, 3947–3956.

[197] Yu, M. J, Thrasher, K. J, McCowan, J. R, Mason, N. R, & Mendelsohn, L. G. Quinazolinone cholecystokinin B receptor ligands. *J. Med. Chem.* (**1991**) 34, 1505–8.

[198] Yu, M. J, McCowan, J. R, Mason, N. R, Deeter, J. B, & Mendelsohn, L. G. Synthesis and X-ray crystallographic analysis of quinazolinone cholecystokinin/gastrin receptor ligands. *J. Med. Chem.* (**1992**) 35, 2534–42.

[199] Varnavas, A, Lassiani, L, Luxich, E, Zacchigna, M, & Boccu, E. Quinazolinone derivatives: synthesis and binding evaluation on cholecystokinin receptors. *Farmaco.* (**1996**) 51, 333–9.

[200] Padia, J. K, Chilvers, H, Daum, P, Pinnock, R, Suman-Chauhan, N, Webdale, L, & Trivedi, B. K. Design and synthesis of novel nonpeptide CCK-B receptor antagonists. *Bioor. Med. Chem. Lett.* (**1997**) 7, 805–10.

[201] Padia, J. K, Field, M, Hinton, J, Meecham, K, Pablo, J, Pinnock, R, Roth, B. D, Singh, L, Suman-Chauhan, N, Trivedi, B. K, & Webdale, L. Novel nonpeptide CCK-B antagonists: design and development of quinazolinone derivatives as potent, selective, and orally active CCK-B antagonists. *J. Med. Chem.* (**1998**) 41, 1042–1049.

[202] Bertrand, P, Bohme, G. A, Durieux, C, Guyon, C, Capet, M, Jeantaud, B, Boudeau, P, Ducos, B, Pendley, C. E, Martin, G. E, & et al. Pharmacological properties of ureidoacetamides, new potent and selective non-peptide CCKB/gastrin receptor antagonists. *Eur. J. Pharmacol.* (**1994**) 262, 233–45.

[203] Howbert, J. J, Lobb, K. L, Britton, T. C, Mason, N. R, & Rasmussen, K. Diphenylpyrazolidinone and benzodiazepine cholecystokinin antagonists: a case of convergent evolution in medicinal chemistry. *Bioor. Med. Chem. Lett.* (**1993**) 3, 875–80.

[204] Castro, J. L, Broughton, H. B, Russell, M. G, Rathbone, D, Watt, A. P, Ball, R. G, Chapman, K. L, Patel, S, Smith, A. J, Marshall, G. R, & Matassa, V. G. 5-(Piperidin-2-yl)- and 5-(homopiperidin-2-yl)-1,4- benzodiazepines: high-affinity, basic ligands for the cholecystokinin B receptor. *J. Med. Chem.* (**1997**) 40, 2491–2501.

[205] Castro, J. L, Ball, R. G, Broughton, H. B, Russell, M. G, Rathbone, D, Watt, A. P, Baker, R, Chapman, K. L, Fletcher, A. E, Patel, S, Smith, A. J, Marshall, G. R, Ryecroft, W, & Matassa, V. G. Controlled modification of acidity in cholecystokinin B receptor antagonists: N-(1,4-benzodiazepin-3-yl)-N'-[3-(tetrazol-5- ylamino) phenyl]ureas. *J. Med. Chem.* (**1996**) 39, 842–849.

[206] Hagishita, S, Seno, K, Kamata, S, Haga, N, Ishihara, Y, Ishikawa, M, & Shimamura, M. Potent and subtype-selective CCK-B/gastrin receptor antagonists: 2,4- dioxo-1,5-benzodiazepines with a plane of symmetry. *Bioorg. Med. Chem.* (**1997**) 5, 1433–1446.

[207] Tokarski, J. S & Hopfinger, A. J. Three-dimensional molecular shape analysis-quantitative structure- activity relationship of a series of cholecystokinin A receptor antagonists. *J. Med. Chem.* (**1994**) 37, 3639–3654.

[208] van der Bent, A, ter Laak, A. M, AP, I, & Soudijn, W. Molecular modelling of asperlicin derived cholecystokinin A receptor antagonists. *Eur. J. Pharmacol.* (**1992**) 226, 327–34.

[209] Sinha, J, Kurup, A, Paleti, A, & Gupta, S. P. Quantitative structure-activity relationship study on some nonpeptidal cholecystokinin antagonists. *Bioorg. Med. Chem.* (**1999**) 7, 1127–30.

[210] Klebe, G & Abraham, U. On the prediction of binding properties of drug molecules by comparative molecular field analysis. *J. Med. Chem.* (**1993**) 36, 70–80.

[211] Kubinyi, H. (**1998**) in *Encyclopedia of Computational Chemistry*, ed. v. Rague Schleyer, P. (John Wiley & Sons).

[212] Klebe, G, Abraham, U, & Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* (**1994**) 37, 4130–46.

[213] Klebe, G & Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput. Aided Mol. Des.* (**1999**) 13, 1–10.

[214] Kearsley, S. K & Smith, G. M. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron. Computer. Methodology.* (**1990**) 3, 615–633.

[215] Bohm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput. Aided Mol. Des.* (**1992**) 6, 593–606.

[216] Bohm, H. J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* (**1992**) 6, 61–78.

[217] Weiner, S. J, Kollman, P. A, Nguyen, D. T, & Case, D. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* (**1986**) 7, 230–252.

[218] Ghose, A. K & Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships i. partition coefficients as a measure of hydrophobicity. *J. Comp. Chem.* (**1986**) 7, 565–577.

[219] Gasteiger, J & Marsili, M. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* (**1980**) 36, 3219–3288.

[220] Klebe, G, Mietzner, T, & Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput. Aided Mol. Des.* (**1999**) 13, 35–49.

[221] Schafferhans, A & Klebe, G. Docking ligands onto Binding Site Representation Derived from Proteins built by Homology Modelling. *J. Mol. Biol.* (**2001**) 307, 407–427.

[222] Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Chem. Ges.* (**1894**) 27, 2984–2993.

[223] Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* (**1994**) 8, 243–56.

[224] Ajay & Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* (**1995**) 38, 4953–67.

[225] Gohlke, H & Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chemie Intern. Ed.* (**2002**) 41, 2644–2676.

[226] Taylor, R. D, Jewsbury, P. J, & Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* (**2002**) 16, 151–166.

[227] Bissantz, C, Folkers, G, & Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* (**2000**) 43, 4759–4767.

[228] Stahl, M & Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* (**2001**) 44, 1035–1042.

[229] McConkey, B. J, v. Sobolev, & Edelman, M. The performance of current methods in igand-protein docking. *Curr. Science* (**2002**) 83, 845–856.

[230] Klebe, G, Mietzner, T, & Weber, F. Different approaches toward an automatic structural alignment of drug molecules: Apllications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput. Aided Mol. Des.* (**1994**) 8, 751–778.

[231] Kelley, L. A, Gardner, S. P, & Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* (**1996**) 9, 1063–5.

[232] Jagerschmidt, A, Guillaume, N, Roques, B. P, & Noble, F. Binding sites and transduction process of the cholecystokininB receptor: involvement of highly conserved aromatic residues of the transmembrane domains evidenced by site-directed mutagenesis. *Mol. Pharmacol.* (**1998**) 53, 878–85.

[233] Beinborn, M, Lee, Y. M, McBride, E. W, Quinn, S. M, & Kopin, A. S. A single amino acid of the cholecystokinin-B/gastrin receptor determines specificity for non-peptide antagonists. *Nature* (**1993**) 362, 348–50.

[234] Gigoux, V, Maigret, B, Escreut, C, Poirot, S, Bouisson, M, Fehrentz, J. A, Moroder, L, Gully, D, Martinez, J, Vaysse, N, & Fourmy, D. Identification of new determinants and interactions of the CCK-A receptor agonist binding site that stabilize its high affinity phopholipase-C coupled state. *In. Press.* (**1998**).

[235] Müller, G, Batoulis, J, Burger, T, & Hessler, G. Monte Carlo based conformational analysis. *Internal Bayer AG report* (**year?**).

[236] Baxter, C. A, Murray, C. W, Clark, D. E, Westhead, D. R, & Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* (**1998**) 33, 367–382.

[237] Schafferhans, A. Entwicklung einer Methode zur Zusammenführung protein- und ligandenbasierter Verfahren zum rationalen Wirkstoffdesign bei unbekannter Rezeptorstruktur. *Thesis Department of Pharmacy of the Philipps University Marburg* (**2000**).

[238] Rarey, M, Kramer, B, Lengauer, T, & Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* (**1996**) 261, 470–489.

[239] Rarey, M, Wefing, S, & Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput. Aided Mol. Des.* (**1996**) 10, 41–54.

[240] Hibert, M. F, Trumpp-Kallmeyer, S, Bruinvels, A, & Hoflack, J. Three-dimensional models of neurotransmitter G-binding protein- coupled receptors. *Mol. Pharmacol.* (**1991**) 40, 8–15.

[241] Kraulis, P. J. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst.* (**1991**) 24, 946–950.

[242] Word, J. M, Lovell, S. C, LaBean, T. H, Taylor, H. C, Zalis, M. E, Presley, B. K, Richardson, J. S, & Richardson, D. C. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* (**1999**) 285, 1711–33.

[243] Romano, R, Dufresne, M, Prost, M. C, Bali, J. P, Bayerl, T. M, & Moroder, L. Peptide hormone-membrane interactions. Intervesicular transfer of lipophilic gastrin derivatives to artificial membranes and their bioactivities. *Biochim. Biophys. Acta.* (**1993**) 1145, 235–42.

[244] Moroder, L, Romano, R, Guba, W, Mierke, D. F, Kessler, H, Delporte, C, Winand, J, & Christophe, J. New evidence for a membrane-bound pathway in hormone receptor binding. *Biochemistry* (**1993**) 32, 13551–13559.

[245] Escherich, A, Lutz, J, Escrieut, C, Fourmy, D, van Neuren, A. S, Müller, G, Schafferhans, A, Klebe, G, & Moroder, L. Peptide/Benzodiazepine hybrids as ligands of CCKA and CCKB Receptors. *Biopolymers* (**2001**) 56, 55–76.

Chapter 9

List of Publications

"This is not the end, it is not even the beginning of the end, but it is perhaps the end of the beginning."

- Sir Winston Leonard Spencer Churchill (**1874-1965**), Britisch Prime Minister and Nobel Prize in Literatur laureate

List of Papers

- Brisson, A., Huetz, P., Ringler, P., van Neuren, S., Kremer, F., Fraaije, H. and Engberts, J. *Functionalized lipid tubules: A tool for helical crystallization of proteins.* Progress in Biophysics & Molecular Biology (**1996**) 65(S1): PD107.

- Huetz, P., van Neuren, S., Ringler, P., Kremer, F., van Breemen, J.F.L., Wagenaar, A., Engberts, J.B.F.N., Fraaije, J.G.E.M. and Brisson, A. *Relationship between molecular structure and supramolecular morphology of DODA-EO2-biotin and related lipids.* Chemistry and Physics of Lipids (**1997**) 89: 15-30.

- van Neuren, A.S., Müller, G., Klebe, G. and Moroder, L. *Molecular modelling studies on G protein-coupled receptors: from sequence to structure?* Journal of Receptor and Signal Transduction Research (**1999**) 19 (1-4): 341-53.

- Escherich, A., Lutz, J., Escrieut, C., Fourmy, D., van Neuren, A.S., Müller, G. and Moroder, L. *Benzodiazepine/Peptide Hybrids as Ligands for CCK-A and CCK-B Receptors.* In: Peptides 1999 (eds.: Bajusz, S., and Hudecz, F.) Akademiaia Kiado, Budapest (**1999**) 80-81.

- Escherich A., Lutz J., Escrieut C., Fourmy D., van Neuren A.S., Müller G., Schafferhans A., Klebe G., Moroder L. *Peptide/benzodiazepine hybrids as ligands of CCK-A and CCK-B receptors.* Biopolymers (**2001**) 56(2): 55-76.

- van Neuren, A.S., Müller, G., Klebe, G. and Moroder, L. *A New Modelling Approach towards GPCR-Ligand Complexes: Exploring Antagonist Binding Site of the CCK-B Receptor.* (**2001**) In Preparation.

List of Oral-Presentations

- van Neuren, A.S., Müller, G., Klebe, G. and Moroder, L. *Molecular modelling studies on G protein-coupled receptors: from sequence to structure?* 8th Swiss Workshop of Methodology in Receptor Research, Morschach - Switzerland (**May 1998**).

- van Neuren, A.S., Schafferhans, A., Müller, G., Klebe, G. and Moroder, L. *A New Modelling Approach towards GPCR-Ligand Complexes: Exploring Antagonist Binding Site of the CCK-B Receptor.* Biophysics & modelling of the 7-trans-membrane receptor gene family, Oxford - United Kingdom (**December 2001**).

- van Neuren, A.S., Stein, M., *Design of a small, rationalized compound library from glycidic scaffolds.* 12th European Carbohydrate Symposium, Grenoble - France (**July 2003**).

List of Poster-Presentations

- van Neuren, A.S., ter Laak, A.M. Nederkoorn, P.H.J., Donne-op den Kelder, G.M., and Timmerman, H. *Three Dimensional Model for the Histamine H_1-receptor.* 2nd Symposium Undergraduate Research, Groningen - the Netherlands (**November 1992**).

- van Neuren, A.S., ter Laak, A.M. Nederkoorn, P.H.J., Donne-op den Kelder, G.M., and Timmerman, H. *Three Dimensional Model for the Histamine H_1-receptor.* PAC-Congress, Leiden - the Netherlands (**March 1993**).

- van Neuren, A.S., ter Laak, A.M. Nederkoorn, P.H.J., Donne-op den Kelder, G.M., and Timmerman, H. *Three Dimensional Model for the Histamine H_1-receptor.* Summer Congress KNCV (Royal Dutch Chemical Society), Rotterdam - the Netherlands (**August 1993**).

- Huetz, P., Ringler, P., van Neuren, A.S., Engberts, J., Wagenaar, A., Fraaije, H. and Brisson, A, *Functionalized Lipid Tubules: Tools for Helical Crystallization of Proteins.* Lunteren Meeting SON - Proteinresearch, Lunteren - the Netherlands (**December 1995**).

- van Neuren, A.S., Huetz, P., Ringler, P., and Brisson, A, *Functionalized Lipid Tubules: Tools for Helical Crystallization of Proteins.* 11th European Congress on Electron Microscopy, Dublin - Ireland (**August 1996**).

- van Neuren, A.S., Huetz, P., Ringler, P., Kremer, F.J.B., van Breemen, J.F.L., Engberts, J.B.F.N., Wagenaar, A., Fraaije, J.G.E.M. and Brisson, A, *Functionalized Lipid Tubules: Tools for Helical Crystallization of Proteins.* GBB Symposium, Groningen - the Netherlands (**September 1996**).

- van Neuren, A.S., van Breemen, J., Keegstra, W. and Brisson, A. *Image Analysis of Frozen-hydrated helical crystals of Streptavidin.* Lunteren Meeting SON - Proteinresearch, Lunteren - the Netherlands (**December 1996**).

- van Neuren, A.S., van Breemen, J.F.L., Wagenaar, A., van Esch, J., Huetz, P., Kremer, F.J.B., Engberts, J.B.F.N., and Brisson, A, *F Tubules of Biotinylated Lipids.* GBB Symposium, Groningen - the Netherlands (**September 1997**).

- Escherich, A., Lutz, J., Escrieut, C., Fourmy, D., van Neuren, A.S., Müller, G. and Moroder, L. *Benzodiazepine/Peptide Hybrids as Ligands for CCK-A and CCK-B/Gastrin Receptors.* European Peptide Symposium, Budapest - Hungary (**September 1998**).

- Escherich, A., Lutz, J., Escrieut, C., Fourmy, D., van Neuren, A.S., Müller, G. and Moroder, L. *Benzodiazepine/Peptide Hybrids as Ligands for CCK-A and CCK-B/Gastrin Receptors.* 4. Deutsches Peptidsymposium, Leipzig - Germany (**March 1999**).

- van Neuren, A.S., Schafferhans, A., Müller, G., Klebe, G. and Moroder, L. *A New Modelling Approach towards GPCR-Ligand Complexes: Exploring Antagonist Binding Site of the CCK-B Receptor.* MGMS 18th International Meeting - Modelling Biomolecular Mechanism, York - United Kingdom (**April 2000**).

- Schafferhans, A., van Neuren, A.S. and Klebe, G. *A New Modelling Approach towards Protein-Ligand Complexes: Methodology.* MGMS 18th International Meeting: Modelling Biomolecular Mechanism, York - United Kingdom (**April 2000**).

- van Neuren, A.S., Schafferhans, A., Müller, G., Klebe, G. and Moroder, L. *A New Modelling Approach towards GPCR-Ligand Complexes: Exploring Antagonist Binding Site*

of the CCK-B Receptor. Biophysics & modelling of the 7-trans-membrane receptor gene family, Oxford - United Kingdom (**December 2001**).

Chapter 10

Acknowledgement

"Kind words can be short and easy to say, but their echoes are truly endless."

- Mother Theresa
  (**1900-1998**), nun and Nobel Peace Prize laureate

I have to thank here everybody who helped me with doing the research described in this thesis, helped me with writting this thesis and helped me with the very important “secondary” things. Several people I have to thank by name.

First of all of course **Dr. Gerhard Müller**. He was a very inspiring and very knowledgeble supervisor, although his speed of correcting this thesis was not as quick as he drives with his german cars. However, he always stayed very positiv: “*Wir schaffen es schon !*”. It took some time, but finally he was right, *we have made it!* The decision of doing my PhD-study at Bayer under his supervision was a decision I will never regret to have made. No other word, sentence or essay than the simple German word *Doktorstiefvater* can better describe the importance and the function of this person for the aspirant doktor.

Simultaneously I have to thank **Prof. Dr. Gerhard Klebe** very much for being my *Doktorvater*, for his patience and help with this thesis and also for the nice and enjoyable visits to his research group in Marburg.

Prof. Dr. Louis Moroder I have to thank for his enthousiastme, his enormous knowledge of the CCK receptors and his help with the research of this thesis.

Many times I have refered in this thesis to the research of **Dr. Andrea Schafferhans.** In this place I not only thank her very much for letting me apply her docking-program (DRAGHOME), but I also thank her very much for the pleasant and good collaboration. The bugs in her initial program nerved me first a little bit, but without her, her inspiring work and her final bugfree program Chapter 6 of this thesis would have never been that good and nice. Furthermore, it was a pleasure to live for three months in Marburg on the same address.

I have to thank **Gudrun Laube** for helping me with the work on neural networks. Sadly enough that nice piece of work did not make it into the final version of this thesis. Not only was it nice working with her, but also the many discussions about everything, especially the German-Dutch items, were nice and instructive. Furthermore, I have to thank her very much of letting me be a photographer at her wedding.

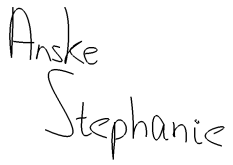
Dr. Gerhard Heßler, Anja Kreis, Svend Matthiesen (even learned a little bit of Dutch) and **Dirk Mahlcke** (burned many CDs) were all nice, good and helpful colleagues. The many discussions with them about scientific and other secondary items were great and very enjoyable and helped to create a great working environment.

Klaudia Augustin did not directly help with the research of this thesis, but I have to thank her very much for the many tennis matches we played against each other. Although I lost many of these matches I enjoyed them a lot. Only that unpredictable backhand I did not like.

Finally of course my parents (**pa & ma**). They have cleaned several houses and packed and also unpacked many movingboxes several times. Without them the moves from Holland to Germany, from Germany to Holland and from Holland back to Germany would have been virtual impossible. They were and still are a great help in many ways. *Dank jullie wel.*

When I left the Netherlands to go to Germany I did not know what to expect and it was a jump in deep water. However, the continuously problems of full hard disks on many computers, the

many computer crashes, the many burned CDs, the enormous amount of trees that went through the printer, and the thickness of this booklet are silent witnesses that during the three years at Bayer not only coffee was cooked. My time in Leverkusen turned out to be more than three very instructive years in which I learned a lot, they were also nice, enjoyable and very pleasant in which I made good and valuable friends. Once more: *thank you all, herzlichen Dank, dank jullie wel !!*



Chapter 11

Curriculum Vitae

"The value of life lies not in the length of days, but in the use of them: a man may live long, yet get little from life. Whether you find satisfaction in life depends not on your tale of years, but on your will."

- Michel de Montaigne
  (**1533-1592**, philosopher and writer)

Personalia

Name :	Anske Stephanie van Neuren
Date of birth :	14 november 1970
Place of birth :	Mijdrecht, the Netherlands
Nationality :	Netherlands

Education

sep. 1977 - jun. 1983	Elementary School, Mijdrecht, the Netherlands.
aug. 1983 - jun. 1989	Secondary School: Athenaeum "Veenlanden College", Mijdrecht, the Netherlands. subjects: Dutch, English, Chemistry, Mathematics A & B, Physics, Biology and Economy

Pharmacochemistry Study

sep. 1989 - aug. 1990	Propedeutical Examination in the Department of Chemistry and Pharmacochemistry of the "Vrije Universiteit" of Amsterdam, the Netherlands.
sep. 1990 - aug. 1994	Doctoral Examination in the Department of Chemistry and Pharmacochemistry of the "Vrije Universiteit" of Amsterdam, the Netherlands.

Contents of the Doctoral examination (Master of Science)

Main Program	Pharmacochemistry & Biochemistry
Basic Courses	Molecular Toxicology, Design/QSAR, Synthesis in Pharmacochemistry, Chemical Computer Science
Examinations	QSAR, General Pharmacochemistry, Computational Chemistry, Chemical Computer Science
Major	Design & Synthesis at the Design-group of the Department of Pharmacochemistry of the "Vrije Universiteit" of Amsterdam, the Netherlands. "The construction of a 3D-model for the Histamine H_1-receptor, a G protein-coupled receptor."
Minor	Chemical Computer Science at the Theoretical Chemistry-group of the Department of Chemistry of the "Vrije Universiteit" of Amsterdam, the Netherlands. "Molecular Electrostatical Potentials (MEPs), how to compute ?"

Industrial Research Project	At the division of Medicinal Chemistry of the pharmaceutical company Solvay Duphar Weesp, the Netherlands "Design of new antagonists of the fibrinogen receptor, glycoprotein IIb-IIIa, which is important for platelet aggregation."
Literature Thesis	"Cell Adhesion Molecules in physiology and pharmacology, with special emphasis on the family of integrin receptors."
Extra	Course to write about Science for the general public.

Post Graduate Research Experiences

okt. 1994 - jun. 1995	Pilot research project at the Design-group of the Department of Pharmacochemistry of the "Vrije Universiteit" of Amsterdam, the Netherlands. "Determination by protein modelling of the similarities between heterotrimeric G-proteins, and other GTP-binding and related proteins with solved 3D-structures."
aug. 1995 - okt. 1997	Research project at the Electron Microscopy group of the Department of Chemistry of the University of Groningen, the Netherlands. "Investigation by electron microscopy of functionalized lipid-tubules: tools for helical crystallization of proteins."
courses	Electron Crystallography, Protein X-ray Crystallography, Gatan DigitalMicrograph and SlowScan CCD Operator's School, Biomolecular Modelling
Masterclass	Membrane Proteins: Reconstruction into Lipid Bilayers and 2D-Crystallization.

PhD-study

nov. 1997 - jan. 2001	A Molecular Modeling study of a G-protein coupled receptor, the human CCK-B/ Gastrin receptor. At the pharmaceutical company Bayer AG, Leverkusen Germany under supervision of Dr. G. Müller (Bayer AG, Germany) and Professor Dr. G. Klebe (Department of Pharmaceutical Chemistry, Philipps University Marburg, Germany).
sep. 1999 - dec. 1999	Research project at the group of Professor Klebe of the Department of Pharmaceutical Chemistry, Philipps University Marburg, Germany. "Docking of Non-peptide CCK-B Antagonists in the Modelled Structure of the Human CCK-B receptor."

Post Doctoral Research Experiences

feb. 2001 - dec. 2002	Molecular Modelling/Cheminformatic position at the Unilever Health Institute - Unilever, Vlaardingen, the Netherlands
jan. 2003 - present	PostDoctoral research position in the field of molecular modelling and computational chemistry at Anterio consult & research GmbH, Mannheim, Germany. "Glycidic Scaffolds: New carbohydrate-derived scaffolds for the generation of bioactive libraries."